

THIS WEEK

EDITORIALS

OCEANOGRAPHY What the good ship Google means for science **p.410**

WORLD VIEW Risky virus research must be halted then discussed **p.411**



DISGUISE How the 390-million-year-old trilobite got its spots **p.412**

Under the gun

A ban on advocacy and promotion of gun control is keeping US agencies from conducting research that is sorely needed to inform policy on firearms and prevent shootings.

US president Barack Obama was expected this week to sign the bill that will fund the US government for the rest of the 2013 fiscal year. The bill contains the first congressional action on gun control since 20 children and six staff members were gunned down at Sandy Hook Elementary School in Newtown, Connecticut, on 14 December. But the congressional action is notable for the wrong reasons. Far from promising to shine badly needed light on the myriad unanswered questions around current policies on firearms, it instead continues restrictions on the gun-related research that is essential to wise policy-making. It is precisely the opposite of what is needed in the United States, where people murder each other with guns at about 20 times the rate in other wealthy nations.

We have been here before. After a gunman murdered 12 people and injured 58 in a Colorado cinema last July, this publication noted that the US National Rifle Association (NRA), through proxies in Congress, has since 1996 inserted restrictive language in the annual bill that funds the Centers for Disease Control and Prevention (CDC). The language prohibits the agency from spending any money “to advocate or promote gun control” (see *Nature* **488**, 129; 2012).

In 2012, that language was extended to include all of the CDC’s huge parent department, the Department of Health and Human Services, which includes the National Institutes of Health (NIH). The biomedical agency had drawn the gun lobby’s ire by funding a study that concluded that guns did not protect their owners from being shot in an assault (C. C. Branas *et al.* *Am. J. Public Health* **99**, 2034–2040; 2009). Admittedly, the study was criticized by some for methodological weaknesses. But NRA chief Wayne LaPierre used it to dismiss all NIH-funded, gun-relevant research as “public health junk science” in a 2010 online commentary. Soon, Congress extended the ban to encompass the NIH.

In the law that Obama signs this week, the NRA-backed prohibition will be continued. Strictly read, it does nothing to bar research, which is distinct from advocacy and promotion, Obama said in January, as he announced his response to the Newtown killings.

That response included, commendably, a memorandum that directed the CDC and other agencies to launch a vigorous effort to study the causes of gun violence and ways to prevent it. Obama asked Congress to provide US\$10 million in new money to the CDC to support its investigation of the most pressing questions, adding that White House lawyers do not interpret such research as contravening the existing prohibition. His administration argued in an accompanying document: “Research on gun violence is not advocacy; it is critical public health research that gives all Americans information they need.”

And they do need it. The 17-year dearth of government-funded research, and a parallel chill on collection of gun-related data by US agencies, have left politicians without the most rudimentary information needed to inform policy-making. Consider, for instance, that the most recent Department of Justice data on the proportion of gun sales occurring between private individuals — as opposed to sales by

licensed dealers — date to 1994. Not since 2004 has the CDC included firearms-related questions in its Behavioral Risk Factor Surveillance System questionnaire, an instrument used to shape state policies. The number of academic publications relevant to firearm violence has fallen by 60% since peaking in 1996, according to an analysis by the coalition Mayors Against Illegal Guns, which looked at trends from 1960 to 2010.

“Obama’s research agenda will not get far in the presence of the existing prohibition.”

But however strong the need, Obama’s research agenda will not get far in the presence of the existing prohibition, which was crafted to intimidate researchers and funding agencies — and has been having precisely that effect.

The irony is that the gun lobby and its congressional allies might benefit from rigorous research. Would a robust study reveal that state laws allowing citizens to carry concealed weapons have resulted in more or fewer deaths? We don’t know. Would the spiking homicide rate in Chicago, Illinois, be higher still if it were not for the city’s restrictive gun laws, or are those laws ineffective? We don’t know. Does a limit on assault weapons reduce the overall rate of firearms injuries and deaths? We don’t know.

How many Newtowns will it take until Congress funds the research that will answer these and other pressing questions? Only strong evidence-based policies would allow politicians to tell grieving parents after the next slaying: we are doing everything we can to stop this happening again. ■

Disciplinary action

How scientists share and reuse information is driven by technology but shaped by discipline.

The words ‘technology’ and ‘revolution’ are being bandied around a lot in scientific publishing — and this week *Nature* presents a special series of articles that explores the industry’s changing landscape (see page 425). But beyond the early adopters of digital technologies who shout freedom from the rooftops and the publishers who look on nervously, what do researchers make of it? Would it surprise you to learn, for example, that more than one-third of academic chemists disagree with the statement “all papers should be published open access”?

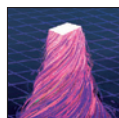
The transformation of research publishing is less a revolution and more a war of attrition. Battle lines were drawn long ago and all sides are well dug-in. In 2001, this journal published a series of viewpoints on the future of ‘e-access to the primary literature’ (see go.nature.com/pezj84).

Those attitudes seem strikingly familiar today. At the time, the founders of the Public Library of Science initiative (then PLS, now reborn as the publisher PLoS) urged that research results should be stored in free, online, centralized repositories. Technology enthusiasts sang the praises of easy search and retrieval across a wide range of publication formats beyond the traditional journal article, but warned of the need for common standards. Publishers pointed out that someone would have to finance the publication of the increasing tide of information, and debated where revenue sources should come from.

There was a voice missing from that debate: yours. More than a decade on, that by surveying more than 23,000 scientists about their experience of and opinions on open-access publishing. The key question is not just what scientists could have, but what they want. The survey results — which NPG plans to release soon — suggest that many scientists are still thinking through their views on the open sharing that the Internet enables, and on whether they want to publish their research openly.

One preliminary result that stands out is the diversity of experiences and attitudes across disciplines. In biology, 17% of papers published by the respondents over the past three years had been immediately made free for all to read by paying the publisher up front, and more than half of the biologists surveyed said that they had published at least one such paper. In chemistry, the proportion of papers was just 4%, and less than one-quarter of chemists had published at least one open-access paper. More than half of biologists felt that “all papers should be published open-access”, whereas just under one-third of chemists agreed (the remaining one-third of chemists neither agreed nor disagreed).

Nor do scientists hold consistent views about how widely



THE FUTURE OF PUBLISHING

A *Nature* special issue.

nature.com/scipublishing

information should be shared and reused. In the NPG survey, 45% felt that all papers should be published open access, but only 22% wanted to allow articles to be reused for commercial purposes. A differently worded survey by publishers

Taylor & Francis of some 14,500 scholars (split between sciences and the arts and humanities) found a similar inconsistency: 40% strongly agreed or agreed that their work should be “reused in any way”, but only 18% said that it was acceptable for others to use their work for commercial gain (W. Frass, J. Cross and V. Gardner *Open Access Survey* Taylor & Francis; 2013). The figures perhaps represent a lack of understanding rather than deeply considered views. For example, when NPG asked scientists which open-access licence they had chosen, including share-alike, no derivatives and CC-BY, 85% of people said: “I don’t know”.

New technologies allow a much greater and faster transition to a digital future, and this week’s special issue reveals that scientists are finding a multitude of ways to publish and access their research results. As this journal has noted before, the future of research literature will ideally be an amalgam of papers, data and software that interlinks with tools for analysis, annotation, visualization and citation. The need for common standards is as great as ever.

But it is demand, not supply, that will shape how scientists and publishers grasp these opportunities. For instance, a key reason that online open-access journals are now accepted as a mainstream (if still minority) method of publishing research is because of the mandates steadily introduced since 2001 by institutions and by research funders.

The dazzling variety of publishing options will fragment the information available on the web. Scholars need to think through how they would like that information to be shared and reused — answers may be different for the various disciplines. One revolution does not yet fit all. ■

Push the boat out

The latest private research vessel to be launched could open up the world of marine science.

The 1984 film *The NeverEnding Story* opens with a bullied boy who forgets his troubles when he reads about an alternative world called Fantasia. So it is tempting to ask whether a new private research vessel, named after one of the characters from the film, can offer hard-pressed oceanographers a similar escape.

As we report on page 420, the R/V *Falkor* has launched into troubled waters for marine science. Ship time is hard to come by, and funds for academic ocean research are shrinking, at least in the United States. More than ever, researchers are needing to cobble together money from several sources to pay for voyages that, in the past, might have been funded through one overarching grant.

Arriving at a rate of knots on this scene are private foundations. The Gordon and Betty Moore Foundation in Palo Alto, California, for instance, has a marine-microbiology initiative that explores sites ranging from coastal upwellings to deep-sea sediments. But it does not accept unsolicited proposals from scientists.

The non-profit Schmidt Ocean Institute, which operates the *Falkor* from its base in Palo Alto, does. It is open to applications to use the ship — for free — for studies that highlight the challenges that the world’s oceans face. The money comes from institute founder Eric Schmidt, former chief executive of Google.

Many researchers are sceptical about the genuine scientific value of millionaire marine philanthropy. It is fashionable to send well-equipped expeditions to little-known parts of the sea, but such efforts seldom connect with the wider scientific community — although film director James Cameron said this week that he would donate his *DEEPSEA*

CHALLENGER submersible to the Woods Hole Oceanographic Institution (WHOI) in Massachusetts. The *Falkor*, however, is meant to be a long-term resource for use by anyone who can get his or her proposal accepted by a peer-review panel.

The ship deliberately has no home port, so as to keep away from territorial claims to one ocean basin or for one country. US and Canadian scientists are scheduled to lead its first few cruises, but that is mainly an artefact of how the Schmidt institute started up. Its programme managers personally visited leading oceanographers in those countries to introduce the *Falkor* and to encourage them to apply for ship time. Non-Americans seem to be taking note; in the latest round of proposals, scientists from 23 countries applied.

The *Falkor* signals the future of oceanography in other ways. It is, at heart, a Google ship, and as such promotes projects that integrate the latest technology into ocean exploration. Some academic outfits, such as the WHOI, already push this limit, but the Google crew has some ideas of its own and scientists will follow them eagerly. The Schmidt institute also prods its shipboard scientists to release their data openly and rapidly, following the strong precedent set by the Moore foundation.

A small ship such as the *Falkor* will not solve all the woes of oceanographers. One possible stumbling block is that the Schmidt institute awards only ship time; scientists must come up with their own funding to pay salaries and for post-cruise science. Thus, only researchers who are accomplished enough to secure funding from other sources will be able to spend time aboard.

The *Falkor* has much to offer, but it will not and should not replace the current US research fleet, which is due to receive much-needed upgrades over the next ten years that will allow it to push the boundaries of marine science further than ever before.

The NeverEnding Story was a joint project of film-makers in Germany and the United States. The *Falkor* was originally a German fishery-protection vessel. With fair seas and a following wind, its story has some way to run yet. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhnuq



H5N1 viral-engineering dangers will not go away

Governments, funders and regulatory authorities must urgently address the risks posed by gain-of-function research, says **Simon Wain-Hobson**.

Barely two months after a small group of influenza virologists lifted a moratorium on work to make the H5N1 avian flu virus as transmissible between humans as seasonal flu, researchers are at it again. Earlier this month, a Dutch scientist proposed similar experiments with other avian flu viruses, as well as the SARS coronavirus. And a fortnight ago, scientists in Germany and Switzerland reported how they had tweaked canine distemper virus to make it grow in human cells.

The logic behind these kinds of experiments, collectively called gain-of-function (GOF) research, is to identify combinations of mutations that could allow an animal virus to jump to unprepared humans. By knowing the mutations, the thinking goes, we can better prepare and marshal our scientific defences against a possible threat.

GOF research on avian flu provoked heated controversy, much of it covered by this journal. That controversy did not go away with the lifting of the moratorium. On the contrary, it continues to fester. Officials in Washington DC are putting the finishing touches to new guidelines for the review, regulation and oversight of this kind of research. The chill winds that we can anticipate blowing from policy-makers as a result could affect all of us who research viruses and their pathology. To avoid this, researchers in this field need to learn lessons from the past.

Rather than use the avian flu moratorium to seek advice, listen and foster debate, many influenza scientists engaged in an academic exercise of self-justification. There was a single large open meeting, at the Royal Society in London, which engaged a wider audience, including bioethicists. The recent calling off of the moratorium by 40 flu researchers alone — not funders, governments or international bodies — says it all. The flu community simply hasn't understood that this is a hot-button issue that will not go away.

There are parallels here to my own field of HIV. In the early days of research, HIV scientists, buoyed by huge research monies, exuded hubris, promised a vaccine within two years, and all sorts of other things. The crunch came when they realized that they had to engage seriously with patient groups. The result is that HIV patients became the most faithful collaborators of HIV clinicians. It is too easy for scientists in a field to dismiss criticism and ideas from outside.

Here are the issues that must be openly addressed about GOF work with avian flu, the SARS coronavirus — or any other virus.

First, is the virological basis for the work sound? The outcomes of the H5N1 experiments are dominated by the artificial-selection systems used. If aerosol-transmitted virus is systematically passed from ferrets with severe respiratory distress, then the research teams will end up with a transmissible and highly virulent strain. Likewise, if animals with mild symptoms

are chosen, a transmissible virus of low virulence would ultimately emerge. Whether nature will take any of these courses is unknown. Take dog breeding. Ruthless selection of alleles over a short period has produced phenomenal phenotypic variation — dachshunds, salukis, whippets and setters. Would nature have come up with the dachshund?

Second, infectious-disease researchers are fond of saying that microbes do not respect barriers. So who makes the rules and provides oversight? Barely a sound has emerged at the international level. The World Health Organization has held essentially closed-door meetings and has failed singularly to widen the debate.

Third, what if these groups generate a highly pathogenic and transmissible virus — which I suspect, within two years, they will? Then what? Should the virus be shared? Should research on this novel virus strain of catastrophic potential be highly restricted?

Fourth, what if there were a leak or a small outbreak? Crippling lawsuits would follow. Are the academic institutions sufficiently covered in terms of insurance? Are university regents or chancellors even aware of the power, and dangers, of the modern molecular biology going on in their labs? Again, not a word has emerged.

Fifth, the world has never been more densely populated. Is it appropriate for civilian scientists to make microbes more dangerous? Is creating a novel human virus antisocial? Was there a failure of duty on the part of funders and regulators? What is the ethical position on such work? Here there has been a start, but as yet there is no consensus.

The global ramifications of GOF research have simply not been sufficiently explored and discussed. Influenza virologists are going down

a blind alley and the powers that be are blindly letting them go down that alley, which is tantamount to acquiescing. So let's be clear: the end game could be viruses more dangerous than the Spanish flu strain.

H5N1 GOF work — indeed all virological GOF work — should be suspended until virologists open up and engage in public discussion of their work and the issues it raises. Given that the flu community failed utterly to use the year-long hiatus to good effect, it is clear that an independent risk-benefit assessment of GOF work is needed.

Governments, funders and regulatory authorities should encourage the influenza virologists to listen and discuss. Learned societies should get off the fence and speak out. A conference involving all the stakeholders is needed, as happened at Asilomar in the 1970s for recombinant DNA. The problem will not go away. It has to be engaged and it has to be done now. ■

Simon Wain-Hobson is chair of the Foundation for Vaccine Research, which is based in Washington DC.
e-mail: simon.wain-hobson@pasteur.fr

**INFLUENZA
VIROLOGISTS ARE
GOING DOWN A
BLIND ALLEY
AND THE POWERS
THAT BE ARE BLINDLY
LETTING THEM GO
DOWN THAT ALLEY.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/g1cqrs

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

BEHAVIOURAL NEUROSCIENCE

Male mice cut out for fathering

Male mice that live with their mates help to nurture pups, whereas virgin males attack them, owing to differences in neural signalling. Kashiko Tachikawa and her colleagues at the RIKEN Brain Science Institute in Saitama, Japan, compared how sexually naive males and cohabiting fathers reacted to pups displaced from their nests, and found that their behaviour was mediated by the pheromone-detecting vomeronasal organ located at the base of the nasal cavity. Most of the virgin males attacked the pups, and the neural pathways that link the vomeronasal organ to other areas in their brains were highly activated in these animals. Cohabiting fathers, however, carried pups back to their nests and displayed little activation in these neural pathways. Sexually naive males showed parental behaviour if their vomeronasal organ had been removed.

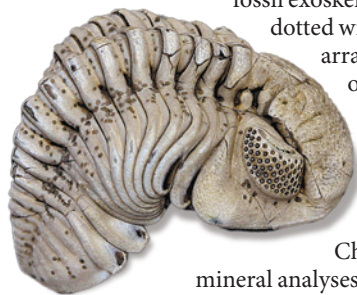
J. Neurosci. 33, 5120–5126 (2013)

PALAEONTOLOGY

Trilobite fossil spotted

Speckles on 390-million-year-old trilobite fossils (pictured) may have helped to camouflage the marine invertebrates.

A team led by Christopher McRoberts, of the State University of New York at Cortland and Thomas Hegna at Western Illinois University



in Macomb examined more than 25 *Eldredgeops rana* trilobites that were discovered in central and western New York state. The

fossil exoskeletons were dotted with regular arrangements

of a variable number of spots — more than 500 in some cases.

Chemical and mineral analyses indicated that spots are of similar composition to the rest of the fossil, suggesting the markings are not artefacts.

No pigment molecules were

preserved, but the authors speculate that the spots may have scattered sunlight and blurred a trilobite's shadow. *Geology* <http://dx.doi.org/10.1130/G34158.1> (2013)

STRUCTURAL BIOLOGY

Serotonin signalling secrets

Researchers have deciphered molecular structures of two neurotransmitter receptors bound with drug molecules.

Bryan Roth at the University of North Carolina Chapel Hill Medical School and Raymond Stevens at the Scripps Research Institute in

many fruit-tree seedlings as the hunted sites. The researchers suggest that with many fewer primates eating fruits and dispersing the seeds by spitting and defaecation, fruit trees do not reproduce as well in hunted areas. As these hunted forests age, the lack of seedlings will mean the dying trees are not replaced, and primates will be unlikely to find enough to eat. *Proc. R. Soc. B* 280, 20130246 (2013)

Primates hunting leaves forest scar

By hunting for primates, humans may create an ecological ripple that makes the hunted regions permanently unsuitable for the animals.

Ola Olsson at Lund University in Sweden and his colleagues looked at six sites in three protected areas in south-eastern Nigeria. Three of the sites were well protected from hunters, and had more than three times as



ECOLOGY

La Jolla, California, and their colleagues, solved the crystal structure of a serotonin receptor called 5-HT_{2B} bound to the migraine medication ergotamine, which is chemically related to the hallucinogen lysergic acid diethylamide (LSD).

In a related paper, a team led by Stevens and Eric Xu at the Van Andel Research Institute in Grand Rapids, Michigan, solved the structure of another serotonin receptor, 5-HT_{1B}, bound to ergotamine. Differences between the structures show how similar drugs can have different biological effects.

TIM LAMAN/NATUREPL.COM

CHRISTOPHER MCROBERTS

Such insight could help to design receptor-specific therapeutics and limit side effects, the authors say.

Science <http://dx.doi.org/10.1126/science.1232807> (2013); *Science* <http://dx.doi.org/10.1126/science.1232808> (2013)

For a longer story on this research, see go.nature.com/yw8pdm

NEUROIMAGING

Single cells seen in whole brains

A fast imaging technique can monitor the activity of individual neurons across the brain of fish larvae, providing both a detailed and comprehensive view of a vertebrate brain.

Misha Ahrens and Philipp Keller at Janelia Farm Research Campus in Ashburn, Virginia, imaged the brains of transparent larval zebrafish (*Danio rerio*), which had been engineered to have neurons that fluoresce when active. The authors improved on a technique that uses planes of laser light to capture images layer by layer, called light-sheet microscopy, by increasing the imaging speed. The method could cover the full brain of a 5-day-old zebrafish larva every 1.3 seconds, detecting more than 80% of its 100,000-odd neurons at the single-cell level (pictured).

Such images can help researchers to map activity

across the brain and show which parts of the brain function together, the researchers say. *Nature Meth.* <http://dx.doi.org/10.1038/nmeth.2434> (2013) For a longer story on this research, see go.nature.com/f8dtkd

OBESITY

Fat cells use immune signal

In obese individuals, fat cells may act like inflammatory white blood cells by using communication machinery once considered to be exclusive to immune cells.

High numbers of several types of immune cell patrol the fat tissues of people who are obese. This inflammation is highly correlated with health risks such as insulin resistance and metabolic syndrome, although how it gets started is unclear. Willa Hsueh and her colleagues at the Methodist Hospital in Houston, Texas, show that it may begin with adipocytes, or fat cells. The authors compared gene expression in lean and obese women and noticed that adipocytes expressed many genes connected to the major histocompatibility complex II, which certain white blood cells use to activate other white blood cells. An array of follow-up studies in mice fed high-fat diets indicate that it is the fat cells themselves that prompt inflammation.

The researchers suggest that obesity triggers alterations in fat cells that are then escalated by white blood cells.

Cell Metab. 17, 411–422 (2013)

DEVELOPMENT

Origins of the ear lining

In mammalian embryos, the middle-ear chamber starts as a pouch that pokes out from the oral cavity. The roof of this cavity ruptures and fills with cells that then retract up over the bones of the middle ear, leaving an air-filled space.

Hannah Thompson and

COMMUNITY CHOICE

The most viewed papers in science

CHEMISTRY

Long DNA-like chains assemble

HIGHLY READ
on www.acs.org
in February

Extremely long molecular chains can self-assemble from two nucleobase-like chemicals, report Nicholas Hud at the Georgia Institute of Technology in

Atlanta, and his co-workers.

The team showed that three of each of the monomers cyanuric acid and a pyrimidine derivative can form a rosette structure. When combined in water, the rosettes stacked themselves into long chains that were more than 1 micrometre in length, the longest self-assembling water-soluble structures ever made from such small molecules. Hydrogen bonding and stacking between the monomers was similar to that seen between individual nucleobases in DNA and RNA. The work provides insight into how a self-assembling system might have existed in a prebiotic world.

J. Am. Chem. Soc. 135, 2447–2450 (2013)

Abigail Tucker at King's College London looked at mice genetically engineered such that the descendants of cells in the developing ear could be tracked. This revealed a double origin for the epithelial cells within the cavity, and thus in two cell types. Cells lining the bottom of the cavity are thick and have cilia that robustly sweep away debris. Cells lining the roof are flat and offer little protection from debris or inflammation. The way in which the mammalian ear develops and our susceptibility to middle-ear infections may be related, the authors suggest.

Science 339, 1453–1456 (2013)

PLANETARY SCIENCE

Migrating planets sped up collisions

The pockmarked asteroid Vesta (pictured) has provided more hints about the history of the early Solar System.

As giant planets such as Jupiter and Saturn shifted into place about 4.1 billion years ago, their changing gravitational pull flung asteroid fragments into highly eccentric orbits. Those space



rocks, in turn, slammed into other asteroids at high speed, heating them and altering their chemistry.

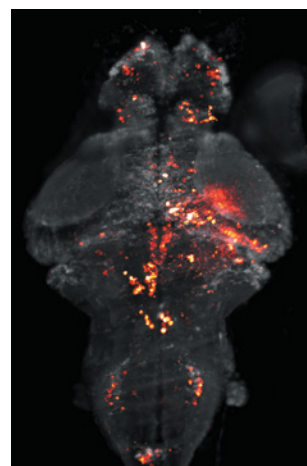
This scenario, proposed by Simone Marchi at the Southwest Research Institute in Boulder, Colorado, and his colleagues, could explain why the asteroid Vesta was pummelled so frequently between 4.1 billion and 3.4 billion years ago. The team's simulations suggest that meteors were slung at much higher velocities than once thought, for a period that lasted hundreds of millions of years.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo1769> (2013)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

FUNDING

US budget boost

Some US science agencies will receive a modest funding boost under government spending plans worth US\$984 billion in total for 2013, as approved by Congress on 21 March. The budgets announced for the National Science Foundation, Food and Drug Administration and the National Institutes of Health provide follow-on finances from current spending plans that expire on 27 March. But the budget increases do not negate across-the-board funding cuts of 5.1% that began on 1 March. See go.nature.com/pvoord for more.

French protest

French scientists went on strike on 21 March to protest against a bill to reform research and higher education. "French universities are being deliberately driven to bankruptcy by laws," campaigners said. Research minister Geneviève Fioraso acknowledged that budget deficits were limiting government spending on science. The bill was adopted by the French cabinet on 20 March and will be presented to parliament on 27 May. It should become law by the end of July. See go.nature.com/xz7sl7 for more.

NUMBER CRUNCH

£4.4 m

The amount (US\$6.7 million) anonymously donated on 19 March to Britain's cash-strapped Royal Institution to save it from having to sell its 214-year-old London home to pay off debts.



BEZOS EXPEDITIONS

Apollo engines rise from the deep

Parts of the Saturn V rocket engines that powered NASA's *Apollo* astronauts to the Moon have been recovered from the floor of the Atlantic Ocean. The recovery team — led by Jeff Bezos, founder of the online company Amazon.com — announced on 20 March that it had hauled up enough pieces to rebuild two F-1

engines, which lifted the rockets off the launch pad and then fell into the sea (pictured). NASA launched 13 Saturn Vs, each of which had five F-1 engines, between 1967 and 1973. "We found so much," says Bezos, "an incredible sculpture garden of twisted F-1 engines that tells the story of a fiery and violent end."

Stem-cell grants

The California Institute for Regenerative Medicine (CIRM) in San Francisco has awarded US\$32 million in grants for research on human induced pluripotent stem (iPS) cell lines. Winners include seven California-based teams that will derive iPS cells from specific tissues to model disease in individual patients. The United States lacks a central iPS-cell bank like those proposed in Japan, but the California initiative will give US researchers access to well-characterized iPS lines for research, says the CIRM.

Canadian budget

Canada's 2013 budget, published on 21 March, has been welcomed by

applied researchers but criticized by basic scientists. It includes Can\$383 million (US\$383 million) for research infrastructures, university research partnerships with industry and support for growing innovative businesses. But James Turk, executive director of the Canadian Association of University Teachers in Ottawa, notes that "there is a consistent pattern of steering money away from basic research". See go.nature.com/5plxzb for more.

POLICY

DNA transplants

The United Kingdom has edged closer to being the first nation to legalize *in vitro* fertilization techniques

that prevent children from inheriting diseases caused by faulty maternal mitochondrial DNA. Legislators in London agreed on 20 March to provide the government with regulatory advice on technologies that transplant nuclear DNA from an egg with diseased mitochondria into a healthy donor cell. A consultation found that there was broad public support for the techniques if they are used to prevent serious diseases. The UK government will now decide whether to legalize the technology. See go.nature.com/gxvwkv for more.

Misconduct rebuttal

The US Department of the Interior (DOI) has rejected allegations brought by

CLIFF MOORE Paul Houser, a former DOI scientific-integrity officer who claimed in 2012 that he was unfairly dismissed for identifying misconduct. The DOI rejected the claims in its first report on alleged misconduct since introducing a scientific-integrity policy in 2011. The report was sent on 19 March to Public Employees for Environmental Responsibility, a watchdog group in Washington DC. See go.nature.com/1jcafh for more.

AWARDS

Physics prize

The 2013 Fundamental Physics Prize was awarded to Alexander Polyakov of Princeton University in New Jersey for his work on the quantum mathematics that underpin the standard model of particle physics. The prize, announced on 20 March and worth US\$3 million, is an annual award established by Russian Internet entrepreneur Yuri Milner in 2012 to honour recent advances in physics. See go.nature.com/a7u4hr for more.

Abel prize winner

Belgian mathematician Pierre Deligne (pictured) has won the Abel Prize, one of the most prestigious awards in mathematics, worth 6 million Norwegian kroner



(US\$1 million). Deligne, who works at the Institute for Advanced Study in Princeton, New Jersey, was awarded the prize on 20 March for his research over the past 50 years connecting algebra and geometry, and for advancing number and representation theory. The prize is awarded annually by the Norwegian Academy of Science and Letters in Oslo and is named after the Norwegian mathematician Niels Henrik Abel. See go.nature.com/psnyu2 for more.

RESEARCH

Neanderthal genome

The sequence of the entire Neanderthal genome has been completed from a toe bone found in a cave in central Siberia. Researchers at the Max Planck Institute for Evolutionary Anthropology in

Leipzig, Germany, announced on 19 March that they had generated a genome sequence 50 times the quality of a draft Neanderthal genome released in 2010 (R. E. Green *et al.* *Science* **328**, 710–722; 2010). The higher quality genome, which is available online (see go.nature.com/o2kd5d), could help researchers to chart the evolutionary histories of Neanderthals, their Denisovan relatives and modern humans more accurately.

Big Bang pictures

The Planck space telescope has produced the most detailed images yet of the residual glow from the Big Bang. Scientists from the European Space Agency unveiled the results, which give clues to the first moments of the Universe, on 21 March. See page 417 and go.nature.com/utdcjm for more.

BUSINESS

Solar bankruptcy

A Chinese subsidiary of the world's largest manufacturer of solar photovoltaic panels announced bankruptcy on 20 March after defaulting on US\$541 million in loans. Suntech Power Holdings was founded in Wuxi, China, in 2001, and led the solar market until fierce competition from 2008 onwards depressed prices. Eight Chinese banks

COMING UP

1 APRIL

UK funding agencies introduce policies that require research to be published as open access. go.nature.com/7lvzll

2–5 APRIL

In Seefeld, Austria, researchers meet to discuss how ecosystems respond to climate variability and extreme weather. go.nature.com/m7jb9n

asked a court to declare the subsidiary, Wuxi Suntech, as insolvent. The subsidiary did not object to the petition.

PEOPLE

Basic science chief

The National Institute of General Medical Sciences at the US National Institutes of Health in Bethesda, Maryland, appointed a new director on 25 March. Jon Lorsch, a biochemist at Johns Hopkins University in Baltimore, Maryland, will head the US\$2.4-billion-a-year basic-research institute, which has been overseen by acting director Judith Greenberg since 2011.

Australian tensions

Australia has its fourth new cabinet minister for science and research in less than 16 months, after prime minister Julia Gillard appointed Craig Emerson to the post on 25 March. The worrying instability is the result of leadership tensions since Gillard deposed former prime minister Kevin Rudd in 2010. On 22 March, former science minister Chris Bowen (a Rudd supporter) resigned after a failed attempt to unseat Gillard. See go.nature.com/jde9dw for more.

➔ **NATURE.COM**

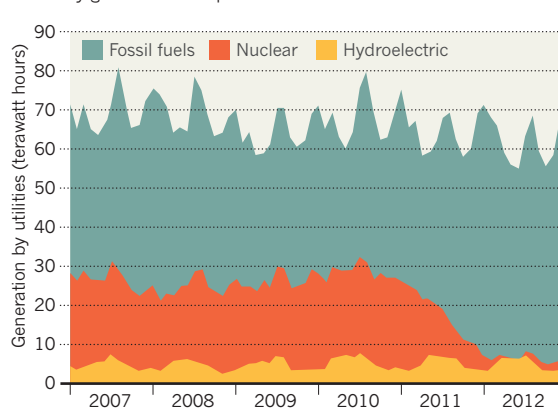
For daily news updates see: www.nature.com/news

TREND WATCH

The amount of electricity generated from fossil fuels rose to 90% of Japan's total electricity output during 2012, according to the US Energy Information Administration (see chart). Only two nuclear reactors are operating two years after the March 2011 earthquake and tsunami that led to three reactor meltdowns at the Fukushima Daiichi nuclear power plant. Japan has imported liquid natural gas and has increased its consumption of crude oil and heavy fuel oil to meet demand.

JAPAN'S FOSSIL-FUEL RELIANCE

Nuclear-plant outages led to a 21% rise in the use of fossil fuels for electricity generation in Japan in 2012.



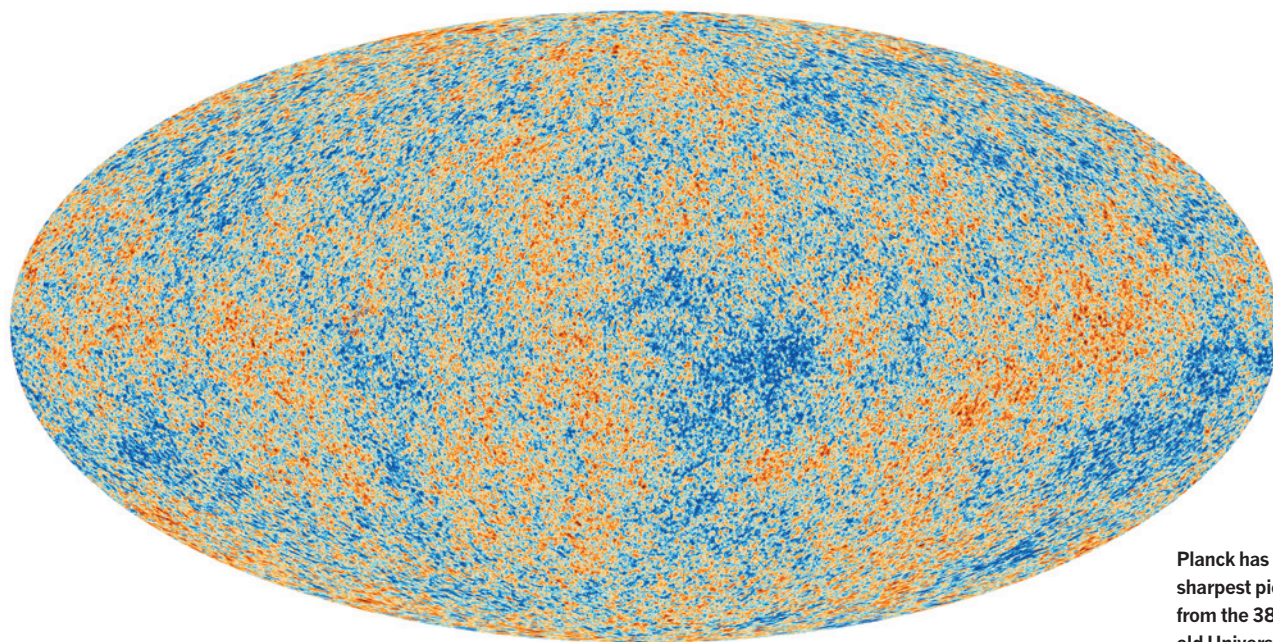
NEWS IN FOCUS

ITALY Minister gives green lights to controversial stem-cell treatments **p.418**

CYBERCRIME Journal impostors snare authors in costly scam **p.421**

PHYSICS How to cook up a Higgs without a massive collider **p.422**

SPECIAL ISSUE *Nature* explores the new landscape of publishing **p.425**



Planck has unveiled the sharpest picture of light from the 380,000-year-old Universe.

COSMOLOGY

Planck snaps infant Universe

Space telescope culls exotic creation theories with ultra-precise microwave map.

BY MARK PEPLow

For astronomers, it is the ultimate treasure map. On 21 March, the Planck space telescope team released the highest-precision map yet of the cosmic microwave background (CMB), the faint but ubiquitous afterglow of the Big Bang. Crowning nearly 50 years of CMB study, the map records the precise contours of the nascent Universe — and in doing so pins down key parameters of the Universe today.

The tiny fluctuations embedded in the CMB map reveal a Universe that is expanding slightly more slowly than had been thought. That dials back the amount of gravity-counteracting 'dark energy' to 68.3% of the Universe, and adds a little more unseen dark matter to the mix. It also means that the Universe is a little older: 13.82 billion years old, adding a few tens of millions of years to the previously calculated value. The map even shows that the number of neutrino 'flavours' permeating the cosmos will probably remain at three — had there been

a fourth, the Universe would have expanded more quickly during its first moments.

These results represent refinements of numbers obtained by previous missions such as the Wilkinson Microwave Anisotropy Probe (WMAP). Where the Planck spacecraft, watching the sky from a vantage point 1.5 million kilometres away, breaks the most new ground is in its support for the reigning theory that describes the instant after the Big Bang. The theory, known as inflation, holds that during an unimaginably rapid expansion lasting just 10^{-32} seconds or so, the Universe grew from a subatomic point to something the size of a grapefruit that then continued to expand at a more stately pace. This growth spurt would help to explain why the Universe we see today

is homogeneous on the largest scales, yet riddled with clumps, filaments and sheets of galaxies.

"Planck could have found that there was

something majorly wrong with inflation," says astrophysicist Jo Dunkley at the University of Oxford, UK, who has worked on data from Planck and the WMAP. "Instead, we've got new evidence that this expansion did happen."

In the minutes that followed the burst of inflation, particles such as protons and electrons formed from the cauldron of proto-matter, and photons began to bounce around like pinballs. It was only 380,000 years later, when the charged plasma cooled into neutral atoms, that those photons could fly freely. Today they make up the CMB, and carry with them an imprint of the quantum fluctuations that roiled the inflationary Universe.

Seen in the map as tiny variations around an average temperature of 2.7 kelvins, the fluctuations caused alterations in the density of matter, which ultimately snowballed into the galaxies seen today. "All the structures we see in the Universe are coming from these little perturbations," says Paul Shellard, a Planck cosmologist at the University of Cambridge, UK. ▶

➔ **NATURE.COM**
For more on the Planck mission:
go.nature.com/ylwzmu

PLANCK COLLABORATION/ESA

► Inflation, proposed in 1980 by Alan Guth, a physicist now at the Massachusetts Institute of Technology in Cambridge, predicts that the range of temperature variations should follow a bell curve — a smooth Gaussian distribution. Cosmologists had found hints in previous missions that the distribution was not so smooth, suggesting that some other process was involved in inflating the Universe (see *Nature* <http://doi.org/bgjd86>; 2008). But so far, Planck's temperature data look almost perfectly Gaussian, and standard theories for inflation are looking ever more secure.

"A lot of baroque inflationary models are gone," says Paul Steinhardt, a theoretical physicist at Princeton University in New Jersey, who has tried to poke holes in inflation by proposing theories such as ekpyrosis, which invokes a cyclical Universe that dies and is reborn in a series of Big Bounces.

But the cosmological case is not quite closed. There are a few details in Planck's map that seem out of place: an odd 'cold patch', for example, and a greater proportion of hotspots on one side of the sky. Moreover, Planck's value for the Hubble constant, which describes the rate of expansion of the Universe, is surprisingly low compared with estimates made with other astronomical techniques — perhaps a hint of new physics in play.

Full confirmation of inflation — and clues about what drove it — will depend on the detailed properties of the CMB's photons. The wrenching moment of inflation should have shaken the very fabric of space-time, resulting in gravitational waves. They in turn may have left a pattern in the polarization of the photons. The Planck team expects to release its polarization data early next year. "If we found gravitational waves, we'd get a Nobel prize — it's a big deal," says George Efstathiou, director of the Kavli Institute for Cosmology in Cambridge, UK, and one of Planck's lead researchers.

The exceedingly faint polarization signal may lie beyond the reach of Planck's detectors. Ground-based microwave telescopes, such as the Keck Array in Antarctica, are also in on the hunt, although they are limited to looking at one hemisphere of the sky, and in certain microwave frequencies, because oxygen in Earth's atmosphere can block some of the CMB photons. Charles Lawrence of NASA's Jet Propulsion Laboratory in Pasadena, California, the lead Planck scientist in the United States, says that it may take another space telescope to finish the job, or perhaps even a mission, decades away, to detect the gravitational waves directly.

But in terms of temperature variations, Lawrence says, astronomers will have to be content with Planck, which "squeezes pretty much all the juice out of the CMB". He finds that juice very sweet, even if it leaves a few questions beyond reach. "We have a pretty good idea of what the Universe is, but we don't have the faintest idea why it is," says Lawrence, adding with impish glee: "It's rather fun, isn't it?" ■



A naked woman joined protesters in Rome calling for stem-cell therapy for all incurably ill patients.

REGENERATIVE MEDICINE

Stem-cell ruling riles researchers

Italian health minister's support for a controversial treatment appals the country's scientists.

BY ALISON ABBOTT

Clinics that offer unproven stem-cell treatments often end up playing cat and mouse with health regulators, no matter which country they operate in. In Italy, however, one such treatment now has official sanction. The country's health minister, Renato Balduzzi, has decreed that a controversial stem-cell treatment can continue in 32 terminally ill patients, mostly children — even though the stem cells involved are not manufactured according to Italy's legal safety standards.

The unexpected decision on 21 March has horrified scientists, who consider the treatment to be dangerous because it has never been rigorously tested. In the opinion of stem-cell researcher Elena Cattaneo of the University of Milan: "It is alchemy".

The decision followed weeks of media pressure to authorize compassionate use of the therapy, which was developed by the Brescia-based Stamina Foundation and has been repeatedly banned in the past six years. Now,

patient groups are pushing for the treatment to be available to anyone with an incurable illness. Hundreds protested in Rome on 23 March, including a naked woman with pro-Stamina slogans painted on her skin.

Stamina Foundation president Davide Vannoni, a psychologist at the University of Udine, says that the publicity around the treatment has won him 9,000 new patients. He hopes that further modifications to the law will allow him to expand the therapy.

A month ago, an investigatory television programme, *The Hyena*, reported that children with incurable diseases such as spinal muscular atrophy were being denied supposedly important treatment, and Italian show-business personalities joined the call to relax rules on stem-cell treatment.

In Italy, the compassionate use of as-yet-unapproved therapies is allowed on an emergency basis for dying individuals who have no other options, and the national health service must provide them for free. The law requires that health authorities approve the quality of such therapies, but some of its terms

are ambiguous, says Amedeo Santosuosso, a Milanese judge and a professor at the University of Pavia who specializes in science and law. “That has been the underlying problem in the Stamina debacle,” he says. “In the case of the Stamina Foundation therapy, there is no suggestion that it might be efficacious, so in my opinion compassionate use is not legitimate.”

Vannoni says that he developed the therapy after having successful stem-cell treatment for a virus-induced facial paralysis in 2004 in Russia. He invited a Russian and a Ukrainian scientist to Turin to develop the method and says that Stamina has since treated 80 or so patients — including people with Parkinson’s disease, Alzheimer’s and muscle-wasting disorders. He has not published the outcomes or precise details of his therapy, which uses the mesenchymal stem cells from bone marrow that differentiate into bone, fat and connective tissue. In his protocol, the cells are extracted from patients, manipulated in the laboratory and then re-infused.

Vannoni acknowledges that he has not published outcomes but says that the method is far from alchemy. Each treatment uses five types of cell, he explains, with their claimed characteristics tuned to replace damaged tissue or to secrete molecules that could reduce inflammation, fight infection or promote blood-vessel growth. “Whatever the disease, one of the types of cell is going to have the right effect,” he says.

When a 2007 European Union regulation required that stem-cell therapies follow the same safety and efficacy rules as pharmaceuticals,

Vannoni moved his lab to the republic of San Marino. “There, rules were not so strict,” he says.

But his work had drawn the attention of a Turin prosecutor, Raffaele Guariniello, whose investigations concluded that Vannoni’s operation could be “dangerous to public health”. Vannoni says that Guariniello marshalled international pressure to stop him working in San Marino, so he moved to Trieste, where he says Guariniello again stopped his work.

From there, Vannoni moved to a public hospital in Brescia. Last May, a delegation including representatives of the Italian Medicines Agency (AIFA) and the ISS, the health ministry’s national institute, visited the Brescia lab and reported chaotic conditions: ethics-committee approvals had been based on inadequate information, and there were no detailed protocols or patient follow-up, for example. The AIFA closed the lab, stating that the facilities could not be trusted to produce contamination-free preparations.

Patients and families turned to the legal system to allow treatments to continue as compassionate use; many of the courts concluded that it was a patient’s right to receive treatment and that health services must offer it, and in some cases the Brescia lab once again supplied cells.

Some of the compelled treatments led to the only publication of clinical results so far. Clinicians at the Burlo Garafalo Children’s Hospital in Trieste treated five babies with type I spinal muscular atrophy and published the results last October (M. Carrozzi *et al.* *Neuromuscul.*

Disord. **22**, 1032–1034; 2012). They found that “the treatment did not change the course of the disease”, says co-author Marco Carrozzi. Vannoni argues that the therapy failed because the clinicians did not use his exact cocktail of cells.

Setting himself against his own regulatory agencies, Balduzzi had earlier angered scientists when, on 7 March, he authorized continued therapy for a three-year-old child with the deadly disease metachromatic leukodystrophy — provided that the stem cells were created in

“The treatment did not change the course of the disease.”

a good manufacturing practice (GMP) facility. Thirteen academics, including Cattaneo and Santosuosso, published an open letter to Balduzzi warning him of the dangers (see go.nature.com/pb1wdl; in Italian).

That authorization was bad enough, says Paolo Bianco, a stem-cell scientist at the University of Rome who co-signed the letter. “Now the minister is allowing the non-GMP version and saying that an unauthorized, unpublished, unknown practice is a ‘treatment.’”

Balduzzi’s decree is likely to be his last legislative act in Italy’s outgoing government, and scientists hope that his successor will respect the role of the AIFA and other science-based agencies. AIFA president Luca Pani declined to comment on the political decision but says that his agency is sticking to its statements on the safety and efficacy of the stem-cell preparations from Brescia. “Our ban holds,” he says. ■

POLICY

Drug-company data vaults to be opened

European agency will publish firms’ clinical-trial results.

BY DANIEL CRESSEY

International calls for the pharmaceutical industry to share the results of clinical trials have grown ever more intense amid revelations that high-profile companies have hidden crucial data on safety and efficacy. Now Europe is moving towards measures that would significantly increase disclosure of such data.

The European Medicines Agency (EMA) in London, which licenses drugs for use in the European Union, is developing a policy to publish some clinical-trial data submitted by companies. And next month, major players in the UK medical community will meet to discuss the practical problems of data openness.

The meeting is likely to take place on

19 April and could feature representatives from biomedical charity the Wellcome Trust, as well as the Academy of Medical Sciences, the Association of the British Pharmaceutical Industry and the Association of Medical Research Charities. It marks a move from discussion to action, says Nicola Perrin, head of policy at the Wellcome Trust in London. “We should all stop discussing whether [the issue is] important or not and start having practical discussions about how we move forward,” she says.

The United States already requires that clinical trials used to secure drug approvals are listed in a public online registry. Other countries have rules that encourage registration. But some researchers and campaigners fear

that key details are not getting into the public domain, making it difficult to assess a drug’s safety and depriving researchers of data.

Critics note, for example, that in recent years the London-based drug company AstraZeneca has been mired in legal problems including accusations that it concealed data on the side effects of its antipsychotic drug Seroquel (quetiapine). And GlaxoSmithKline (GSK), also based in London, has paid out billions of dollars after pleading guilty to charges including misrepresenting the safety of its diabetes drug Avandia (rosiglitazone) and the effectiveness of its antidepressant Paxil (paroxetine) in children. The Cochrane Collaboration, an international group of medical experts, has called for Roche in Basel, Switzerland, to open up its data on the influenza drug Tamiflu (oseltamivir) so that the group can assess the drug’s efficacy.

Drug companies are making concessions. GSK and Roche have agreed to share data with scientists whose requests are deemed legitimate. But critics say that Roche has not established a fully independent group to assess requests.

Last year, following pressure from campaigners, the EMA said that it would proactively publish certain data submitted by drug companies. The agency is consulting industry and academic researchers and funders on ►

► how this would work; it is likely to have a policy in place by 1 January 2014.

“Europe is now ahead of the United States in this area,” says Lisa Bero, a pharmacologist at the University of California, San Francisco, who studies bias in scientific publication. “We’re all waiting”, she adds, to see what will unfold. “This is all very new.”

Bero says that the move towards more transparency is a victory. But “there’s going to be a lot more battles” over how information should be released, how detailed it should be, who should control its release and who should have access to data that might reveal identities of trial participants.

Campaigners such as All Trials, an international group launched in January to put pressure on governments and regulators, have demanded that public registration not be limited to trials used in drug-approval applications and should include full methods and results of past and current trials.

Many medical researchers would be satisfied if summary results were made public, ideally as peer-reviewed papers. But some want full release of the detailed clinical study reports that drug companies submit to regulatory agencies. Researchers who conduct meta-analyses by pooling data from multiple trials also want access to anonymized data on individual participants.

In 2010, the European Ombudsman ruled that detailed data should not be seen as commercially confidential. That helped to drive development of the EMA’s policy, which will increase access to such information. But patient-level data are unlikely to be placed wholesale into the public domain, and companies, researchers, trial funders and patient groups disagree on who should decide which researchers can access them.

Groups also wrestle over who will enforce the rules. The UK Health Research Authority, which oversees standards and regulations in medical research, is assessing the practicalities of assigning the job to the same ethics committees that approve clinical trials.

Kay Dickersin, director of the Center for Clinical Trials at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, says that serious legal clout is the only way to enforce transparency. Fines are not enough, she adds: for serious infringements, “someone has to go to jail”. ■



The R/V *Falkor* can deploy remotely operated submersibles.

OCEANOGRAPHY

Private research ship makes waves

Falkor plots a fresh course for ocean studies with Google cash.

BY ALEXANDRA WITZE

Life onboard the R/V *Falkor* is a far cry from the spartan existence endured by most academic oceanographers on research trips. The privately run research vessel features a sauna, a glassed-in lounge and a helicopter pad to be kept clear at all times for VIP guests.

All of which isn’t too surprising, as two of the VIPs are the *Falkor*’s benefactors: Google board chair Eric Schmidt and his wife Wendy. The 83-metre former fishery-protection vessel was retrofitted by the Schmidt Ocean Institute of Palo Alto, California, at a cost of

US\$60 million. Supporters say that it adds a new capability to US oceanography’s fraying infrastructure: an ability to take on riskier projects more quickly than federal agencies can.

“Our hope is to accelerate the pace of research,” says Wendy Schmidt. “It’s not science as usual.” The ship’s first full-length science cruise ends on 29 March when it returns to port in St Petersburg, Florida, after a three-week trip.

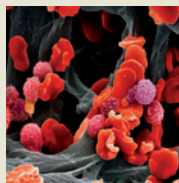
For some oceanographers, the arrival of the *Falkor* — named after a lucky dragon in the film *The NeverEnding Story* — is a stroke of luck. The US research fleet is ageing and shrinking: down from 26 ships in the University-National Oceanographic Laboratory System in 1995 to

DEBBIE NAIL MEYER/SCHMIDT OCEAN INST.



**MORE
ONLINE**

TOP STORY

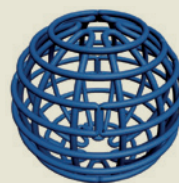


Genetic engineering coaxes immune T cells into fighting leukaemia
go.nature.com/skwai

MORE NEWS

- US writers are less reserved than British ones go.nature.com/ugpp8a
- Lava eruptions led to dinosaur reign go.nature.com/nm6sxl
- Cosmic-ray fluctuations don’t mean Voyager 1 ‘left the Solar System’ go.nature.com/dzwmub

VIDEO



DNA origami uses crosslinks to enter into third dimension
go.nature.com/ziblu4

HAO YAN LAB, BIODESIGN INST.

STEVE G. SCHMEISSNER/SPL

19 today. Federal funds for ship time are dwindling as well. For the first time in 15 years, for example, microbiologist Julie Huber of the Marine Biological Laboratory in Woods Hole, Massachusetts, is relying entirely on private money for her studies. A slot on the *Falkor* in September will allow her to test a microbial sampler on the sea floor — technology development that she says would have progressed more slowly with incremental federal grants.

There is a long history of millionaires dabbling in oceanography without doing much publicly available science. Hollywood director James Cameron has a small group of researchers advising him on his deep-sea dives, such as his record-breaking trip last year to the Mariana Trench. He is donating the submersible used on that dive to the Woods Hole Oceanographic Institution (WHOI) in Massachusetts, although he is holding back the discoveries from that plunge for use in a feature film.

The Schmidt Ocean Institute, established in 2009, aims to reach a wider array of scientists. A parallel set-up, the Marine Science and Technology Foundation, also headed by Eric Schmidt, supports the development of oceanography tools, such as a video recorder for plankton and a surface-water analyser powered by wind and solar energy. “They

are accelerating change in a field which was already changing gradually,” says Kim Juniper, an oceanographer at the University of Victoria in Canada, who will lead a *Falkor* cruise around Vancouver Island in August to study waters with low oxygen levels.

The *Falkor* can also carry equipment from other organizations. In 2014, for instance, it will take the WHOI’s deep-diving Nereus robot to the Mariana Trench.

Given the ship’s Internet ties, it is no surprise that managers favour projects emphasizing open data. Charles Paull, a marine geologist at the Monterey Bay Aquarium Research Institute in Moss Landing, California, is leading the current cruise in the Gulf of Mexico, where his team is mapping an underwater limestone cliff that bears marks from the meteorite impact 65 million years ago that is thought to have killed the dinosaurs off. Paull plans to post his maps on Google Earth within two months — not the two years that most oceanographic data take to trickle out.

Data sharing is not a requirement for research proposals, says Victor Zykov, Schmidt Ocean’s director of research, but those that include an open-data component are ranked higher than those of equal scientific value that do not.

Falkor cruises are scheduled after the

proposals have been peer reviewed by independent scientists. For the round of cruises beginning in 2014, the 48 applications made were whittled down to seven projects. Sixty-one applications have arrived for the cruises beginning in 2015, with a final selection of seven or so expected late this summer, says Zykov. “It’s hard to deny that interest is growing,” he says. Wendy Schmidt adds that if the cruises turn out good science, the institute may consider adding a second ship in the future.

One catch is that although ship time is provided for free, researchers must find a way to fund their salaries and any post-cruise science. For some this isn’t too much of a problem. Chris German at the WHOI, for instance, used leftover NASA money from an earlier cruise to pay for his portion of a *Falkor* trip this summer to the Caribbean, where he will use Nereus to hunt for new hydrothermal vents.

Just about the only other US ship allowed such unfettered exploration is the US National Oceanic and Atmospheric Administration’s *Okeanos Explorer*, German says. But the *Falkor*, with fewer education and outreach obligations, can operate with more focus. “It’s a new way of doing business,” he says. “The model is theoretically good, but how it’s going to work out remains to be seen.” ■ **SEE EDITORIAL P.410**

PUBLISHING

Sham journals scam authors

Con artists are stealing the identities of real journals to cheat scientists out of publishing fees.

BY DECLAN BUTLER

Scientific publishing, meet cybercrime. Two reputable European science journals have fallen prey to identity theft by criminals who have created counterfeit journal websites. These online doppelgängers have duped hundreds of researchers into paying author fees, with the ill-won gains being funnelled to Armenia.

Editors of the victim journals first learned of the scam last year, but their attempts to put a stop to it have so far come to nothing. The crooked websites are masquerading as *Archives des Sciences*, a multidisciplinary journal founded in 1791 and published by the Society of Physics and Natural History of Geneva (SPHN) in Switzerland; and *Wulfenia*, a botany journal published by the Regional Museum of Carinthia in Klagenfurt, Austria.

The scammers attend to the closest of details, displaying on multiple websites not only the titles of the authentic journals, but also their impact factors, postal addresses and

international standard serial numbers — the unique codes used to identify journals.

Editors of the authentic publications fear that the ruse has tainted the reputations of their journals.

“Victims are regularly contacting me to ask about the status of their papers: they transfer money and don’t see their papers published,” says Roland Eberwein, editor-in-chief of the authentic *Wulfenia* and head of the Botanic Center at the Carinthia museum, which includes a herbarium of more than 200,000 specimens.

“We are currently wasting our time trying to fight these people,” says Robert Degli Agosti, editor-in-chief of *Archives des Sciences* and a plant biologist and electrophysiologist at the University of Geneva.

Neither of the authentic journals has its own dedicated website, making them easy prey for

imposters. In response to the scam, however, the SPHN and the Carinthia museum have put warning notices on their home pages, and *Wulfenia* has started publishing its back issues online.

The forged sites look so convincing that they initially misled Thomson Reuters, a metrics company based in New York that produces the Scientific Citation Index and compiles journal impact factors.

But by May last year, the company had become suspicious, writing to the SPHN for an explanation of the “huge discrepancy” between the content of articles in print issues of *Archives des Sciences* — which Thomson Reuters indexes — and on the website. It noted, too, a discrepancy in publishing frequency: “We receive and index 2 issues of each volume for each year, while the website is now listing 12 issues per volume, one each month,” it wrote to the society.

One of the imposters had even persuaded Thomson Reuters to include a link to the false journal in its list of indexed publications; the company moved swiftly to ►



► remove the link when the scam was uncovered. The action triggered “a barrage of complaints and requests to reactivate the link from representatives of the false journal”, says Marie McVeigh, director of content selection at Thomson Reuters. She says that the company has also received enquiries from customers “asking why the articles that had been accepted by one of the false journals were not appearing in our indexes”.

“The quality and integrity of our content is of the greatest importance to us,” adds McVeigh.

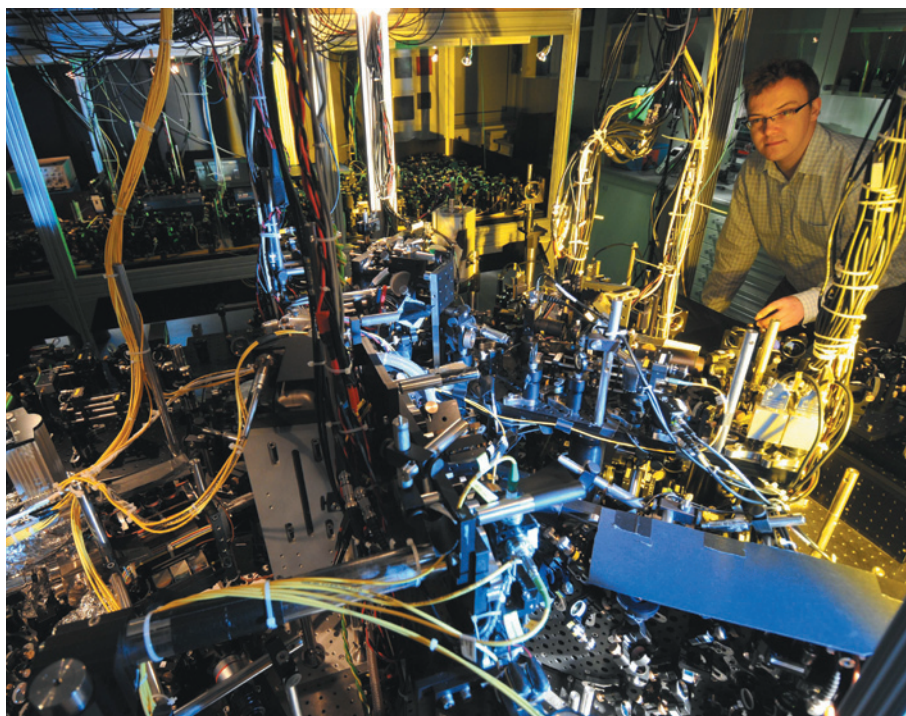
In a further impudent touch, the various counterfeit *Archives des Sciences* websites list an editorial board with 87 members, including Daniel Gamelin, a chemist and materials scientist at the University of Washington in Seattle, and Gerald Cleaver, a high-energy physicist at Baylor University in Waco, Texas. Both are perplexed — and annoyed. “This is the first I have heard of this website or of my listing; I have no affiliation with this organization, nor have I ever,” says Gamelin. Cleaver, too, says that his name is being used without his permission.

The ‘editor-in-chief’ of the fake *Archives des Sciences* journal is named on the counterfeit websites as “Prof. Dr. Eliana Schmid”, with the affiliation “Geneva, Switzerland”. The counterfeit *Wulfenia* sites give as the editor-in-chief Vienna S. Franz and list 35 editorial-board members, with most affiliations giving only city and country. Eberwein and Degli Agosti think that these named editors-in-chief are fictional.

Researchers who have submitted to the fake journals pay dearly. Both the counterfeit *Archives des Sciences* and the fake *Wulfenia* charge author fees of more than \$500, with instructions to address payment to accounts at two banks in Yerevan, Armenia.

Degli Agosti has reported the counterfeit *Archives des Sciences* websites to the Cyber-crime Coordination Unit Switzerland, but was told that the sites were hosted in the United States, so the unit could not act directly against them. However, it advised him that he could press criminal charges under Swiss cybercrime laws. The University of Geneva’s lawyers are helping the SPHN to draft its case, but the society is not part of the university, so it will have to take its case forward alone.

Austrian police have made little progress on the *Wulfenia* case, says Eberwein: they closed down a fake site hosted in Austria, but multiple replicas popped up on servers in other countries. Austrian authorities have told Eberwein that the scope for legal redress is limited, but, unconvinced, he has contacted *Archives des Sciences* to swap notes. ■ **SEE NEWS FEATURE P.433**



THORSTEN NAESE, MAX PLANCK INSTITUTE FOR QUANTUM OPTICS

Physicists have found Higgs-like particles in a superfluid at the Max Planck Institute in Munich, Germany.

PHYSICS

Higgs physics on the cheap

Tabletop Higgs particles may illuminate cosmic cousin.

BY EUGENIE SAMUEL REICH

More than three decades ago, before the world’s most powerful particle collider was even on the drawing board, two physicists discovered a Higgs boson. On a tabletop.

In 1981, Peter Littlewood and Chandra Varma, two solid-state theorists at Bell Laboratories in Murray Hill, New Jersey, realized that a mysterious effect seen in a niobium selenide superconductor could be explained by the jiggling of the invisible field that causes electrons in the material to pair up and move as one without resistance. Mathematically, the disturbance in the field looked very like one that is associated with the Higgs particle found by particle physicists.

Because the superconducting field was already known, the search for the associated particle never drew as much attention as the hunt by particle physicists for the Higgs. That quest culminated on 13 March at the Large Hadron Collider (LHC) at CERN, Europe’s main particle-physics facility near Geneva

in Switzerland, with the announcement of a definitive Higgs detection: the first direct evidence of a Higgs field permeating the Universe and giving objects mass.

But physicists agree that the superconducting Higgs is closely related to its particle-physics cousin. Both arise from the vibration of an invisible field that forces ordinary particles to oscillate in sync. Now physicists are seeing signs that other condensed-matter systems can generate Higgs-like particles, raising hopes that work on one Higgs, studied cheaply on tabletops, can inform the study of another, probed by a US\$5-billion collider. “I’m hoping there will be cross-fertilization,” says Varma, now at the University of California, Riverside, who spoke in a packed session on solid-state Higgs particles at a meeting of the American Physical Society (APS) in Baltimore, Maryland, on 19 March.

➔ **NATURE.COM**
For CERN’s discovery of the Higgs, see:
go.nature.com/gmp3ef

It wouldn’t be the first time that particle physics stood on the shoulder of condensed-matter physics. When Peter

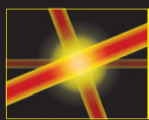
HIGGS HUNTING

Physicists are looking for connections between the cosmic Higgs boson, discovered in a particle collider, and its tabletop cousins.



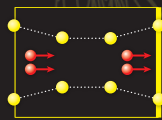
PARTICLE COLLIDER

Energy scale: 1.25×10^{11} eV
Permeates the Universe and gives rise to mass in other particles.



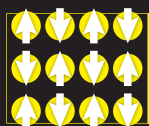
BOSE-EINSTEIN CONDENSATE

Energy scale: 4×10^{-13} eV
Exists as a jiggling in the field describing the shared quantum state of a cloud of atoms.



SUPERCONDUCTOR

Energy scale: 0.002 eV
Exists as a jiggling in the field describing how superconducting electrons pair up.



ANTIFERROMAGNET

Energy scale: Up to 0.0012 eV
Exists as a jiggling in the magnetic ordering of atomic spin states.

eV, electronvolt.

Higgs, a theorist at the University of Edinburgh, UK, put together the idea known as the Higgs field in 1964, he built on a theory developed a year earlier, by theorist Phil Anderson, now at Princeton University in New Jersey, to describe the interactions of superconducting electrons. Higgs took the idea further, interpreting the field as a medium in empty space that would pull on particles, giving them mass. The Higgs bosons that made up the field would be invisible unless the field was jiggled in the right way, as occurs in high-energy collisions at the LHC.

Varma and Littlewood identified their superconducting Higgs in the same way, by recognizing that a wave of slightly displaced atoms was jiggling the superconducting field. Now researchers are trying to extend their catalogue of Higgs particles. At the APS meeting, physicist Marie-Aude Méasson from the University of Paris Diderot and her colleagues reported that the mechanism by which the Higgs was recognized in niobium selenide did not exist in a related superconductor, niobium sulphide, which lacks the wave of displaced atoms. Méasson says that she wants to find examples of other superconductors that can produce Higgs particles. “When you only have one example in nature you wonder if it’s really what you think,” she says.

Other researchers are looking for examples outside superconductors. Christian Rüegg of the Paul Scherrer Institute in Villigen, Switzerland, says that his research group has produced Higgs-like particles in an antiferromagnet, in which the spins of atoms are opposite to adjacent ones. And physicist Immanuel Bloch of the Ludwig Maximilians University in Munich, Germany, told scientists at the APS meeting that he has created Higgs particles in a cloud of chilled rubidium atoms that share the same quantum state (a Bose-Einstein condensate). Using criss-crossing laser beams, Bloch tunes the cloud to a transition point between the atoms’ insulating and superfluid states at which the Higgs particle emerges. Although the speed of light is a factor in equations describing the particle-physics Higgs, it is replaced by the speed of sound in Bloch’s system, which

means that the energy scale of the particles is vastly smaller than that of that of the cosmic Higgs (see ‘Higgs hunting’). This mini-Higgs, says Bloch, is “a scaled-down version”.

Research on such solid-state Higgs particles might illuminate future directions for particle physics, says Subir Sachdev, a physicist at Harvard University in Cambridge, Massachusetts, who organized the APS session. A major disappointment of the Higgs discovered at the LHC is its apparently ‘plain vanilla’ nature. Theorists had hoped for a composite Higgs that would have pointed to the existence of other undiscovered particles and forces. But so far the Higgs has behaved exactly as expected, revealing little about what might lie beyond the standard model of particle physics, thought to be incomplete because it doesn’t include the force of gravity. One obstacle at the LHC is that any undiscovered particles might decay almost immediately after they are created, too quickly for them to be studied. But Bloch’s version of a Higgs particle lasts for one six-hundredth of a second, which allows him to closely observe details of particle decays.

Particle theorist John Ellis of CERN agrees that the study of solid-state Higgs particles might shed light on what lies beyond the standard model. But he notes that ideas can flow in the other direction, too: the mathematical tools developed by particle physicists to study strong interactions between particles are already being used by condensed-matter physicists in the hope of explaining high-temperature superconductors, in which some think that yet more Higgs particles might be lurking. “Maybe particle physics has given something back,” says Ellis. ■

CORRECTION

The News Feature ‘Campus 2.0’ (*Nature* **495**, 160–163; 2013) wrongly said that Andrew Ng went public with Daphne Koller’s machine-learning course — it was his course. Both this course and a database course each attracted 100,000 students not 60,000 as stated.

A NEW PAGE

A special issue of Nature looks at the transformation taking place in scientific publishing.

BRENDAN MONROE

After 350 years in the slow-moving world of print, scientific publishing has been thrust into a fast-paced online realm of cloud computing and ubiquitous sharing. The result has been an era of ferment, as established practices are challenged by new ones — most notably, the open-access model in which the author pays publication fees upfront. Last month, US President Barack Obama's administration declared that government-funded research papers should be made freely available within 12 months of publication (see *Nature* **494**, 414–415; 2013). And from 1 April, research councils in the United Kingdom will require the results of government-funded research to be open access on publication.

In this special issue, *Nature* explores the changing landscape. A News Feature weighs claims that online, author-pays publishing can drastically cut costs (see page 426). Several authors discuss the nuts and bolts of making open-access publishing work well — including copyright pioneer John Wilbanks on open licensing agreements (see pages 440 and 442). A report explores the dark side of open access: publishers whose tactics lead authors to feel disgruntled or duped (see page 433). And a Careers Feature offers advice for researchers trying to balance prestige, cost and career implications in deciding where to submit manuscripts (see page 539).

The special also looks at broader aspects of publishing. Information scientist Jason Priem describes how the concepts of journal and article are being superseded by algorithms that filter, rate and disseminate scholarship as it happens (see page 437). A News Feature investigates how some university libraries are reinventing themselves to help scientists to archive and make accessible a new kind of publication: data sets (see page 430). And Robert Darnton, director of the library at Harvard University in Cambridge, Massachusetts, talks about the soon-to-be launched Digital Public Library of America, which could ultimately hold 5 million books (see page 447). Science itself is changing rapidly; the means by which it is shared must keep up. ■



THE FUTURE OF PUBLISHING

A *Nature* special issue.
nature.com/scipublishing



THE TRUE COST OF SCIENCE PUBLISHING

Cheap open-access journals raise questions about the value publishers add for their money.

BY RICHARD VAN NOORDEN

Michael Eisen doesn't hold back when invited to vent. "It's still ludicrous how much it costs to publish research — let alone what we pay," he declares. The biggest travesty, he says, is that the scientific community carries out peer review — a major part of scholarly publishing — for free, yet subscription-journal publishers charge billions of dollars per year, all told, for scientists to read the final product. "It's a ridiculous transaction," he says.

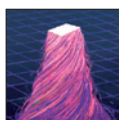
Eisen, a molecular biologist at the University of California, Berkeley, argues that scientists can get much better value by publishing in open-access journals, which make articles free for everyone to read and which recoup their costs by charging authors or funders. Among the best-known examples are

journals published by the Public Library of Science (PLOS), which Eisen co-founded in 2000. "The costs of research publishing can be much lower than people think," agrees Peter Binfield, co-founder of one of the newest open-access journals, *PeerJ*, and formerly a publisher at PLOS.

But publishers of subscription journals insist that such views are misguided — born of a failure to appreciate the value they add to the papers they publish, and to the research community as a whole.

They say that their commercial operations are in fact quite efficient, so that if a switch to open-access publishing led scientists to drive down fees by choosing cheaper journals, it would undermine important values such as editorial quality.

These charges and counter-charges have



THE FUTURE OF PUBLISHING
A *Nature* special issue.
nature.com/scipublishing

BRENDAN MONROE

been volleyed back and forth since the open-access idea emerged in the 1990s, but because the industry's finances are largely mysterious, evidence to back up either side has been lacking. Although journal list prices have been rising faster than inflation, the prices that campus libraries actually pay to buy journals are generally hidden by the non-disclosure agreements that they sign. And the true costs that publishers incur to produce their journals are not widely known.

The past few years have seen a change, however. The number of open-access journals has risen steadily, in part because of funders' views that papers based on publicly funded research should be free for anyone to read. By 2011, 11% of the world's articles were being published in fully open-access journals¹ (see 'The rise of open access'). Suddenly, scientists can compare between different publishing prices. A paper that costs US\$5,000 for an author to publish in *Cell Reports*, for example, might cost just \$1,350 to publish in *PLoS ONE* — whereas *PeerJ* offers to publish an unlimited number of papers per author for a one-time fee of \$299. "For the first time, the author can evaluate the service that they're getting for the fee they're paying," says Heather Joseph, executive director of the Scholarly Publishing and Academic Resources Coalition in Washington DC.

The variance in prices is leading everyone involved to question the academic publishing establishment as never before. For researchers and funders, the issue is how much of their scant resources need to be spent on publishing, and what form that publishing will take. For publishers, it is whether their current business models are sustainable — and whether highly selective, expensive journals can survive and prosper in an open-access world.

THE COST OF PUBLISHING

Data from the consulting firm Outsell in Burlingame, California, suggest that the science-publishing industry generated \$9.4 billion in revenue in 2011 and published around 1.8 million English-language articles — an average revenue per article of roughly \$5,000. Analysts estimate profit margins at 20–30% for the industry, so the average cost to the publisher of producing an article is likely to be around \$3,500–4,000.

Most open-access publishers charge fees that are much lower than the industry's average revenue, although there is a wide scatter between journals. The largest open-access publishers — BioMed Central and *PLoS* — charge \$1,350–2,250 to publish peer-reviewed articles in many of their journals, although their most selective offerings charge \$2,700–2,900. In a survey published last year², economists Bo-Christer Björk of the Hanken School of Economics in Helsinki and David Solomon of Michigan State University in East Lansing looked at 100,697 articles published in 1,370 fee-charging open-access journals active in 2010 (about 40% of the fully open-access articles in that year), and found that charges ranged from \$8 to \$3,900. Higher charges tend to be found in 'hybrid' journals, in which publishers offer to make individual articles free in a publication that is otherwise paywalled (see 'Price of prestige'). Outsell estimates that the average per-article charge for open-access publishers in 2011 was \$660.

Although these fees seem refreshingly transparent, they are not the only way that open-access publishers can make money. As Outsell notes, the \$660 average, for example, does not represent the real revenue collected per paper: it includes papers published at discounted or waived fees, and does not count cash from the membership schemes that some open-access publishers run in addition to charging for articles. Frequently, small open-access publishers are also subsidized, with universities or societies covering the costs of server hosting, computers and building space. That explains why many journals say that they can offer open access for nothing. One example is *Acta Palaeontologica Polonica*, a respected open-access palaeontology journal, the costs of which are

mostly covered by government subsidies to the Institute of Paleobiology of the Polish Academy of Sciences in Warsaw; it charges nothing for papers under 10 pages. Another is *eLife*, which is covered by grants from the Wellcome Trust in London; the Max Planck Society in Munich, Germany; and the Howard Hughes Medical Institute in Chevy Chase, Maryland. And some publishers use sets of journals to cross-subsidize each other: for example, *PLoS Biology* and *PLoS Medicine* receive subsidy from *PLoS ONE*, says Damian Pattinson, editorial director at *PLoS ONE*.

Neither *PLoS* nor BioMed Central would discuss actual costs

(although both organizations are profitable as a whole), but some emerging players who did reveal them for this article say that their real internal costs are extremely low. Paul Peters, president of the Open Access Scholarly Publishing Association and chief strategy officer at the open-access publisher Hindawi in Cairo, says that last year, his group published 22,000 articles at a cost of \$290 per article. Brian Hole, founder and director of the researcher-led Ubiquity Press in London, says that average costs are £200 (US\$300). And Binfield says that *PeerJ*'s costs are in the "low hundreds of dollars" per article.

The picture is also mixed for subscription publishers, many of which generate revenue from a variety of sources — libraries, advertisers, commercial subscribers, author charges, reprint orders and cross-subsidies from more profitable journals. But they are even less transparent about their costs than their open-access counterparts. Most declined to reveal prices or costs when interviewed for this article.

The few numbers that are available show that costs vary widely in this sector, too. For example, Diane Sullenberger, executive editor for *Proceedings of the National Academy of Sciences* in Washington DC, says that the journal would need to charge about \$3,700 per paper to cover costs if it went open-access. But Philip Campbell, editor-in-chief of *Nature*, estimates his journal's internal costs at £20,000–30,000 (\$30,000–40,000) per paper. Many publishers say they cannot estimate what their per-paper costs are because article publishing is entangled with other activities. (*Science*, for example, says that it cannot break down its per-paper costs; and that subscriptions also pay for activities of the journal's society, the American Association for the Advancement of Science in Washington DC.)

Scientists pondering why some publishers run more expensive outfits than others often point to profit margins. Reliable numbers are hard to come by: Wiley, for example, used to report 40% in profits from its science, technology and mathematics (STM) publishing division before tax, but its 2013 accounts noted that allocating to science publishing a proportion of 'shared services' — costs of distribution, technology, building rents and electricity rates — would halve the reported profits. Elsevier's reported margins are 37%, but financial analysts estimate them at 40–50% for the STM publishing division before tax. (*Nature* says that it will not disclose information on margins.) Profits can be made on the open-access side too: Hindawi made 50% profit on the articles it published last year, says Peters.

Commercial publishers are widely acknowledged to make larger profits than organizations run by academic institutions. A 2008 study by London-based Cambridge Economic Policy Associates estimated margins at 20% for society publishers, 25% for university publishers and 35% for commercial publishers³. This is an irritant for many researchers, says Deborah Shorley, scholarly communications adviser at Imperial College London — not so much because commercial profits are larger, but because the money goes to shareholders rather than being ploughed back into science or education.

Commercial publishers are widely acknowledged to make larger profits than organizations run by academic institutions. A 2008 study by London-based Cambridge Economic Policy Associates estimated margins at 20% for society publishers, 25% for university publishers and 35% for commercial publishers³. This is an irritant for many researchers, says Deborah Shorley, scholarly communications adviser at Imperial College London — not so much because commercial profits are larger, but because the money goes to shareholders rather than being ploughed back into science or education.

But the difference in profit margins explains only a small part of the variance in per-paper prices. One reason that open-access publishers have lower costs is simply that they are newer, and publish entirely online, so they don't have to do print runs or set up subscription

**"THE COSTS
OF RESEARCH
PUBLISHING CAN BE
MUCH LOWER THAN
PEOPLE THINK."**

THE COST OF PUBLISHING

JOURNAL PRICES VARY WITH INFLUENCE AND BUSINESS MODEL.

Price of prestige

Open-access prices correlate weakly with the average influence of a journal's articles.

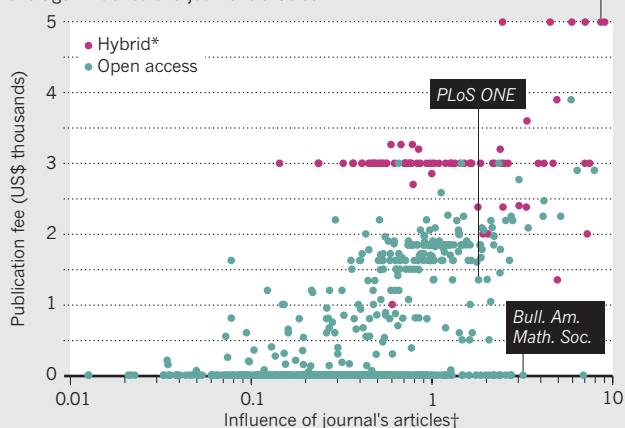
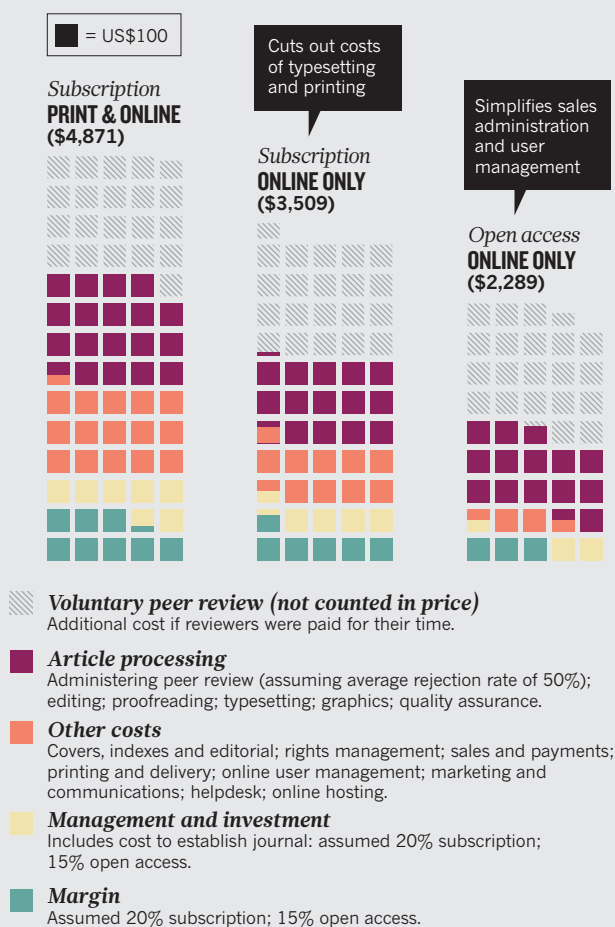


Chart omits open-access journals yet to receive an Article Influence® score.

*Subscription journals that give option of open-access publishing. †Relative score, in which 1 = global average. See www.eigenfactor.org/openaccess for details on how this metric is calculated.

How costs break down

An economic model shows how switching from subscription to open access changes the costs of publishing.



Data from J. Houghton et al. *Economic implications of alternative scholarly publishing models* (Joint Information Systems Committee, 2009), available at go.nature.com/uqrqww.

paywalls (see 'How costs break down'). Whereas small start-ups can come up with fresh workflows using the latest electronic tools, some established publishers are still dealing with antiquated workflows for arranging peer review, typesetting, file-format conversion and other chores. Still, most older publishers are investing heavily in technology, and should catch up eventually.

COSTLY FUNCTIONS

The publishers of expensive journals give two other explanations for their high costs, although both have come under heavy fire from advocates of cheaper business models: they do more and they tend to be more selective. The more effort a publisher invests in each paper, and the more articles a journal rejects after peer review, the more costly is each accepted article to publish.

Publishers may administer the peer-review process, which includes activities such as finding peer reviewers, evaluating the assessments and checking manuscripts for plagiarism. They may edit the articles, which includes proofreading, typesetting, adding graphics, turning the file into standard formats such as XML and adding metadata to agreed industry standards. And they may distribute print copies and host journals online. Some subscription journals have a large staff of full-time editors, designers and computer specialists. But not every publisher ticks all the boxes on this list, puts in the same effort or hires costly professional staff for all these activities. For example, most of *PLoS ONE*'s editors are working scientists, and the journal does not perform functions such as copy-editing. Some journals, including *Nature*, also generate additional content for readers, such as editorials, commentary articles and journalism (including the article you are reading). "We get positive feedback about our editorial process, so in our experience, many scientists do understand and appreciate the value that this adds to their paper," says David Hoole, marketing director at Nature Publishing Group.

The key question is whether the extra effort adds useful value, says Timothy Gowers, a mathematician at the University of Cambridge, UK, who last year led a revolt against Elsevier (see *Nature* <http://doi.org/kwd; 2012>). Would scientists' appreciation for subscription journals hold up if costs were paid for by the authors, rather than spread among subscribers? "If you see it from the perspective of the publisher, you may feel quite hurt," says Gowers. "You may feel that a lot of work you put in is not really appreciated by scientists. The real question is whether that work is needed, and that's much less obvious."

Many researchers in fields such as mathematics, high-energy physics and computer science do not think it is. They post pre- and post-reviewed versions of their work on servers such as arXiv — an operation that costs some \$800,000 a year to keep going, or about \$10 per article. Under a scheme of free open-access 'Episciences' journals proposed by some mathematicians this January, researchers would organize their own system of community peer review and host research on arXiv, making it open for all at minimal cost (see *Nature* <http://doi.org/kwg; 2013>).

These approaches suit communities that have a culture of sharing preprints, and that either produce theoretical work or see high scrutiny of their experimental work — so it is effectively peer reviewed before it even gets submitted to a publisher. But they find less support elsewhere — in the highly competitive biomedical fields, for instance, researchers tend not to publish preprints for fear of being scooped and they place more value on formal (journal-based) peer review. "If we have learned anything in the open-access movement, it's that not all scientific communities are created the same: one size doesn't fit all," says Joseph.

THE VALUE OF REJECTION

Tied into the varying costs of journals is the number of articles that they reject. *PLoS ONE* (which charges authors \$1,350) publishes 70% of submitted articles, whereas *Physical Review Letters* (a hybrid journal that has an optional open-access charge of \$2,700) publishes fewer than 35%; *Nature* published just 8% in 2011.

The connection between price and selectivity reflects the fact

that journals have functions that go beyond just publishing articles, points out John Houghton, an economist at Victoria University in Melbourne, Australia. By rejecting papers at the peer-review stage on grounds other than scientific validity, and so guiding the papers into the most appropriate journals, publishers filter the literature and provide signals of prestige to guide readers' attention. Such guidance is essential for researchers struggling to identify which of the millions of articles published each year are worth looking at, publishers argue — and the cost includes this service.

A more-expensive, more-selective journal should, in principle, generate greater prestige and impact. Yet in the open-access world, the higher-charging journals don't reliably command the greatest citation-based influence, argues Jevin West, a biologist at the University of Washington in Seattle. Earlier this year, West released a free tool that researchers can use to evaluate the cost-effectiveness of open-access journals (see *Nature* <http://doi.org/kwh>; 2013).

And to Eisen, the idea that research is filtered into branded journals before it is published is not a feature but a bug: a wasteful hangover from the days of print. Rather than guiding articles into journal 'buckets', he suggests, they could be filtered after publication using metrics such as downloads and citations, which focus not on the antiquated journal, but on the article itself (see page 437).

Alicia Wise, from Elsevier, doubts that this could replace the current system: "I don't think it's appropriate to say that filtering and selection should only be done by the research community after publication," she says. She argues that the brands, and accompanying filters, that publishers create by selective peer review add real value, and would be missed if removed entirely.

PLoS ONE supporters have a ready answer: start by making any core text that passes peer review for scientific validity alone open to everyone; if scientists do miss the guidance of selective peer review, then they can use recommendation tools and filters (perhaps even commercial ones) to organize the literature — but at least the costs will not be baked into pre-publication charges.

These arguments, Houghton says, are a reminder that publishers, researchers, libraries and funders exist in a complex, interdependent system. His analyses, and those by Cambridge Economic Policy Associates, suggest that converting the entire publishing system to open access would be worthwhile even if per-article-costs remained the same — simply because of the time that researchers would save when trying to access or read papers that were no longer lodged behind paywalls.

THE PATH TO OPEN ACCESS

But a total conversion will be slow in coming, because scientists still have every economic incentive to submit their papers to high-prestige subscription journals. The subscriptions tend to be paid for by campus libraries, and few individual scientists see the costs directly. From their perspective, publication is effectively free.

Of course, many researchers have been swayed by the ethical argument, made so forcefully by open-access advocates, that publicly funded research should be freely available to everyone. Another important reason that open-access journals have made headway is that libraries are maxed out on their budgets, says Mark McCabe, an economist at the University of Michigan in Ann Arbor. With no more library cash available to spend on subscriptions, adopting an open-access model was the only way for fresh journals to break into the market. New

funding-agency mandates for immediate open access could speed the progress of open-access journals. But even then the economics of the industry remain unclear. Low article charges are likely to rise if more-selective journals choose to go open access. And some publishers warn that shifting the entire system to open access would also increase prices because journals would need to claim all their revenue from upfront payments, rather than from a variety of sources, such as secondary rights. "I've worked with medical journals where the revenue stream from secondary rights varies from less than 1% to as much as one-third of total revenue," says David Crotty of Oxford University Press, UK.

Some publishers may manage to lock in higher prices for their premium products, or, following the successful example of PLoS, large open-access publishers may try to cross-subsidize high-prestige,

selective, costly journals with cheaper, high-throughput journals. Publishers who put out a small number of articles in a few mid-range journals may be in trouble under the open-access model if they cannot quickly reduce costs. "In the end," says Wim van der Stelt, executive vice president at Springer in Dordrecht, the Netherlands, "the price is set by what the market wants to pay for it."

In theory, an open-access market could drive down costs by encouraging authors to weigh the value of what they get against what they pay. But that might not happen: instead, funders and libraries may end up paying the costs of open-access publication in place of scientists — to simplify the accounting and maintain freedom of choice for academics. Joseph says that some institutional libraries are already joining publisher membership schemes in which they buy a number of free or discounted articles for their researchers. She worries that such behaviour might reduce the author's awareness of the price being paid to publish — and thus the incentive to bring costs down.

And although many see a switch to open access as inevitable, the transition will be gradual. In the United Kingdom, portions of grant money are being spent on open access, but libraries still need to pay for research published in subscription journals. In the meantime, some scientists are urging their colleagues to deposit any manuscripts they publish in subscription journals in free online repositories. More than 60% of journals already allow authors to self-archive content that has been peer-reviewed and accepted for publication, says Stevan Harnad, a veteran open-access campaigner and cognitive scientist at the University of Quebec in Montreal, Canada. Most of the others ask authors to wait for a time (say, a year), before they archive their papers. However, the vast majority of authors don't self-archive their manuscripts unless prompted by university or funder mandates.

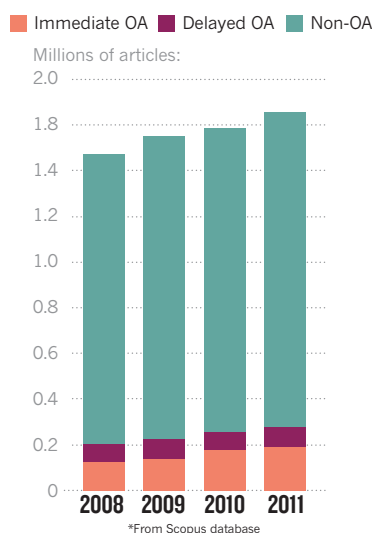
As that lack of enthusiasm demonstrates, the fundamental force driving the speed of the move towards full open access is what researchers — and research funders — want. Eisen says that although PLoS has become a success story — publishing 26,000 papers last year — it didn't catalyse the industry to change in the way that he had hoped. "I didn't expect publishers to give up their profits, but my frustration lies primarily with leaders of the science community for not recognizing that open access is a perfectly viable way to do publishing," he says. ■

Richard Van Noorden is an assistant news editor at *Nature*.

1. Laakso, M. & Björk, B.-C. *BMC Medicine* **10**, 124 (2012).
2. Solomon, D. J. & Björk, B.-C. *J. Am. Soc. Inf. Sci. Technol.* **63**, 1485–1495 (2012).
3. Cambridge Economic Policy Associates *Activities, costs and funding flows in the scholarly communications system in the UK* (Research Information Network, 2008).

THE RISE OF OPEN ACCESS

Immediate open access (OA) made up 12% of articles in 2011*.





THE DARK SIDE OF PUBLISHING

The explosion in open-access publishing has fuelled the rise of questionable operators.

Spam e-mails changed the life of Jeffrey Beall. It was 2008, and Beall, an academic librarian and a researcher at the University of Colorado in Denver, started to notice an increasing flow of messages from new journals soliciting him to submit articles or join their editorial boards. “I immediately became fascinated because most of the e-mails contained numerous grammatical errors,” Beall says. He started browsing the journals’ websites, and was soon convinced that many of the journals and their publishers were not quite what they claimed. The names often sounded grand — adjectives such as ‘world’, ‘global’ and ‘international’ were common — but some sites looked amateurish or gave little information about the organization behind them.

Since then, Beall has become a relentless watchdog for what he describes as “potential, possible or probable predatory scholarly open-access publishers”, listing and scrutinizing them on his blog, Scholarly Open Access. Open-access publishers often collect fees from authors to pay for peer review, editing and website maintenance. Beall asserts that the goal of predatory open-access publishers is to exploit this model by charging the fee

BY DECLAN BUTLER

without providing all the expected publishing services. These publishers, Beall says, typically display “an intention to deceive authors and readers, and a lack of transparency in their operations and processes”.

Beall says that he regularly receives e-mails from researchers unhappy about their experiences with some open-access journals. Some say that they thought their papers had been poorly peer reviewed or not peer reviewed at all, or that they found themselves listed as members of editorial boards they had not agreed to serve on. Others feel they were not informed clearly, when submitting papers to publishers, that publication would entail a fee — only to face an invoice after the paper had been accepted. According to Beall, whose list now includes more than 300 publishers, collectively issuing thousands of journals, the problem is getting worse. “2012 was basically the year of the predatory publisher; that was when they really exploded,” says Beall. He estimates that such outfits publish 5–10% of all open-access articles.

Beall’s list and blog are widely read by librarians, researchers and open-access advocates, many of whom applaud his efforts to reveal shady publishing practices — ones that, they



THE FUTURE OF PUBLISHING
A *Nature* special issue.
nature.com/scipublishing

worry, could taint the entire open-access movement. “I think Beall has taken a brave and principled stand in publishing this, at no small risk to himself,” says Douglas Sipp, an expert in science policy and ethics at the RIKEN Center for Developmental Biology in Kobe, Japan, who studies the open-access movement in Asia.

Beall says that he has been the target of vicious online comments, and last December he was the subject of an online campaign to create the false impression that he was extorting fees from publishers to re-evaluate their status on his list. The Canadian Center of Science and Education, a company based in Toronto that publishes many open-access journals and is on Beall's list, is now threatening to sue him for alleged defamation and libel. But even some experts in scholarly publishing are uncomfortable with Beall's blacklist, arguing that it runs the risk of lumping publishers that are questionable together with those that could be bona fide start-ups simply lacking experience in the publishing industry. Matthew Cockerill, managing director of BioMed Central, an open-access publisher based in London, says that Beall's list “identifies publishers which Beall has concerns about. These concerns may or may not be justified.”

RISING TIDE

As a research librarian, Beall has been in prime position to watch the dramatic changes that have taken place in scientific publishing since the rise of the open-access movement about a decade ago. In the conventional subscription-based model, journals bring in revenue largely through selling print or web subscriptions and keeping most online content locked behind a paywall. But in the most popular model of open access, publishers charge an upfront ‘author fee’ to cover costs — and to turn a profit, in the case of commercial publishers — then make the papers freely available online, immediately on publication.

The open-access movement has spawned many successful, well-respected operations. *PLOS ONE*, for example, which charges a fee of US\$1,350 for authors in middle- and high-income countries, has seen the number of articles it publishes leap from 138 in 2006 to 23,464 last year, making it the world's largest scientific journal. The movement has also garnered growing political support. In the past year, the UK and US governments, as well as the European Commission, have thrown their weight behind some form of open-access publishing. And scarcely a week goes by without the appearance of new author-pays, open-access publishers, launching single journals or large fleets of them.

Many new open-access publishers are trustworthy. But not all. Anyone with a spare afternoon and a little computing savvy can launch an impressive-looking journal website and e-mail invitations to scientists to join editorial boards or submit papers for a fee. The challenge for researchers, and for Beall, is to work out when those websites or e-mail blasts signal a credible publisher and when they come from operations



Jeffrey Beall investigates potentially shady practices in open-access publishing.

“2012 WAS THE YEAR OF THE PREDATORY PUBLISHER; THAT WAS WHEN THEY REALLY EXPLODED.”

that can range from the outright criminal to the merely amateurish.

In one e-mail that Beall received and shared with *Nature*, a dental researcher wrote that she had submitted a paper to an open-access journal after she “was won over by the logos of affiliated databases on the home page and seemingly prestigious editorial board”. But the researcher, who prefers to remain anonymous, says that she became concerned about the peer-review process when the article was accepted within days and she was not sent any reviewers' comments. She says that last week — several months after her original submission — she was sent page proofs that match the submitted manuscript, and that she still has not seen reviewers' comments.

Complaints like this prompted Beall to coin the term predatory publisher and to compile his first list of them, which he published in 2010. He now estimates that his zeal for investigating publishers takes up 20–25 hours a week, squeezed in around his day job. Beall says that he is motivated

partly by his sense of duty, as an academic librarian, to evaluate online resources and to help patrons to “recognize scholarly publishing scams and avoid them”, and partly by the “private and very positive feedback” he receives from researchers and librarians.

But Beall's critics assert that he often relies heavily on analysis of publishers' websites rather than detailed discussions with publishers, and that this might lead to incorrect or premature conclusions. “One of the major weaknesses of Jeffrey Beall's methodology is that he does not typically engage in direct communication with the journals that he has classified as predatory,” says Paul Peters, chief strategy officer at Hindawi Publishing Corporation, based in Cairo, and president of the Open Access Scholarly Publishers Association (OASPA), based

in The Hague, the Netherlands. A set of Hindawi's journals appeared on a version of Beall's list because he had concerns about their editorial process, but has since been removed. “I reanalysed it and determined that it did not belong on the list,” he says. “It was always a borderline case.”

Another concern, say Beall's critics, is that he risks throwing undue suspicion on start-up publishers. “Although rapid launches of many journals may well correlate negatively with journal quality, it is certainly not enough in and of itself to warrant describing a publisher as predatory,” says Cockerill, who is also a board member of the OASPA. “Similarly, some publishers identified on Beall's list are guilty of poor copy-editing and user-interface design on their websites,” he says. “Again, this is, at best, circumstantial evidence for problems with the scholarly standard of the material they publish.”

OMICS Group, based in Hyderabad, India, is on Beall's list. One researcher complained in an e-mail to Beall that she had submitted a paper to an OMICS journal after receiving an e-mail solicitation — but learned that she had to pay a fee to publish it only from a message sent by the journal after the paper had been accepted. “To my horror,

MATT NAGER PHOTOGRAPHY

I opened the file to find an invoice for \$2,700!" she wrote. "This fee was not mentioned anywhere obvious at the time I submitted my manuscript." (*Nature* was unable to contact this researcher.) Beall says that OMICS journals do not show their author fees prominently enough on their journal websites or in e-mails that they send to authors to solicit manuscript submissions.

Srinubabu Gedela, director of OMICS Group, says that article-handling fees are displayed clearly on the 'Instructions for Authors' web page for each OMICS journal. Gedela adds that he would assume researchers would be aware that such open-access journals charge author fees. He says that OMICS Group is "not predatory" and that its staff and editors are acting in "good faith and confidence" to promote open-access publishing.

Publishers in developing countries and emerging economies are at particular risk of being unfairly tarred by Beall's brush, critics say.

Many open-access publishers are springing up in India and China, for example, where swelling researcher ranks are creating large publishing markets. Pressure to publish is often intense in developing countries, and vanity presses could attract unscrupulous researchers keen to pad out their CVs. But respectable domestic publishers could have an important role by helping to address local science issues, such as those related to crops, diseases or environmental problems.

"It is important that criteria for evaluating publishers and journals do not discriminate [against] publishers and journals from other parts of the world," says Lars Bjørnshauge, managing director of the Directory of Open Access Journals (DOAJ), based in Copenhagen, which lists open-access journals that have been reviewed for quality. New publishing outfits may legitimately use aggressive marketing tactics to recruit authors, and they may have yet to polish their websites, editorial boards and peer-review procedures.

"Some are embarrassingly, toe-cringingly amateurish, but predatory is a term that, I think, implies intent to deceive," says Jan Velterop, a former science publisher at Nature Publishing Group and elsewhere who is now working with several start-ups to promote innovative ways to publish science data. Damage could be done if "a damning verdict is given to otherwise honest, though perhaps amateurish, attempts to enter the publishing market", he says.

QUESTIONING THE VERDICT

For researchers involved in journals whose publishers have appeared on Beall's blacklist, the verdict can be unsettling. David Warhurst, a malaria researcher at the London School of Hygiene and Tropical Medicine, is the unpaid editor-in-chief of *Malaria Chemotherapy, Control & Elimination*, an open-access, peer-reviewed journal owned by Ashdin Publishing, a company that has offices in Cairo and Honnelles, Belgium, and that is on Beall's list.

Warhurst says that he was initially reluctant to take up the invitation he received two years ago to become the journal's editor-in-chief, because he found that the publishers did not have a journal registered with the publication index PubMed. But "certainly I do not believe that this is a toxic journal", he says. The journal is still in its launch phase, and refereeing of

BUYER BEWARE

A checklist to identify reputable publishers

How to perform due diligence before submitting to a journal or publisher.

- Check that the publisher provides full, verifiable contact information, including address, on the journal site. Be cautious of those that provide only web contact forms.
- Check that a journal's editorial board lists recognized experts with full affiliations. Contact some of them and ask about their experience with the journal or publisher.
- Check that the journal prominently displays its policy for author fees.
- Be wary of e-mail invitations to submit to journals or to become editorial board members.
- Read some of the journal's published articles and assess their quality. Contact past authors to ask about their experience.
- Check that a journal's peer-review process is clearly described and try to confirm that a claimed impact factor is correct.
- Find out whether the journal is a member of an industry association that vets its members, such as the Directory of Open Access Journals (www.doaj.org) or the Open Access Scholarly Publishers Association (www.oaspa.org).
- Use common sense, as you would when shopping online: if something looks fishy, proceed with caution. **D.B.**

the papers so far has entailed extensive corrections but has not been "exceptional" compared with his experiences at other journals, Warhurst says. The papers "had new findings or findings useful in their geographical context, but needed help with presentation — mainly language and analysis." Ashry Aly, director of Ashdin Publishing, says that the company is not a predatory publisher.

Beall accepts that the publishers on his list fall along a spectrum, with some being worse than others, but he strongly defends his methods. He denies that he doesn't make sufficient efforts to contact publishers, arguing that many of them — who often can be contacted only through a web form — never respond. When it comes to publishers in developing countries, he says: "Look, when I discover a new publisher from Nigeria, I admit I am more suspicious than I would be were the publisher from, for example, the Vatican." But, he says, "I try to be

as fair and honest as I can be when I am making judgements".

Beall says that he usually gives blog posts a 'cooling-off' period between writing and publishing them. Last month, he announced an appeals process in which a three-person advisory board will conduct a blinded review and then recommend whether the publisher or journal should stay on the list. And to improve transparency, Beall last August posted a set of criteria that he says he uses to assess publishers, including an evaluation of their content and practices based on standards established by organizations such as the OASPA. Rick Anderson, a library dean at the University of Utah in Salt Lake City, says that Beall's criteria "make a lot of sense" and also allow for distinctions between truly exploitative publishers and those that are just sloppy.

Bjørnshauge feels that the entire problem needs to be kept in perspective. He estimates that questionable publishing probably accounts for fewer than 1% of all author-pays, open-access papers — a proportion far lower than Beall's estimate of 5–10%. Instead of relying on blacklists, Bjørnshauge argues, open-access associations such as the DOAJ and the OASPA should adopt more responsibility for policing publishers. He says that they should lay out a set of criteria that publishers and journals must comply with to win a place on a 'white list' indicating that they are trustworthy. The DOAJ, he says, is now compiling a list of new, more stringent criteria. To help clean up practices, he adds, research funders should pay author fees only to such white-listed publishers. Meanwhile, he urges researchers to be as cautious when shopping online for publishers as when shopping for anything else (see 'Buyer beware'). "Examine the company you are about to deal with," he says.

Beall says that he supports such efforts. "If someone can figure out a better way, that would be great, and I will defer to them," he says. "I wish them success." But he is sceptical about whether a white list would be able to keep up with the surge of new publishers, and believes that his blacklist provides more immediate warning. That, however, depends on whether he can keep up. "I did not expect and was unprepared for the exponential growth of questionable publishers that has occurred in the past two years," he says. ■

Declan Butler is a senior reporter for *Nature* based in France.

COMMENT

PUBLISHING Proposed new copyright licences could hobble open access **p.440**

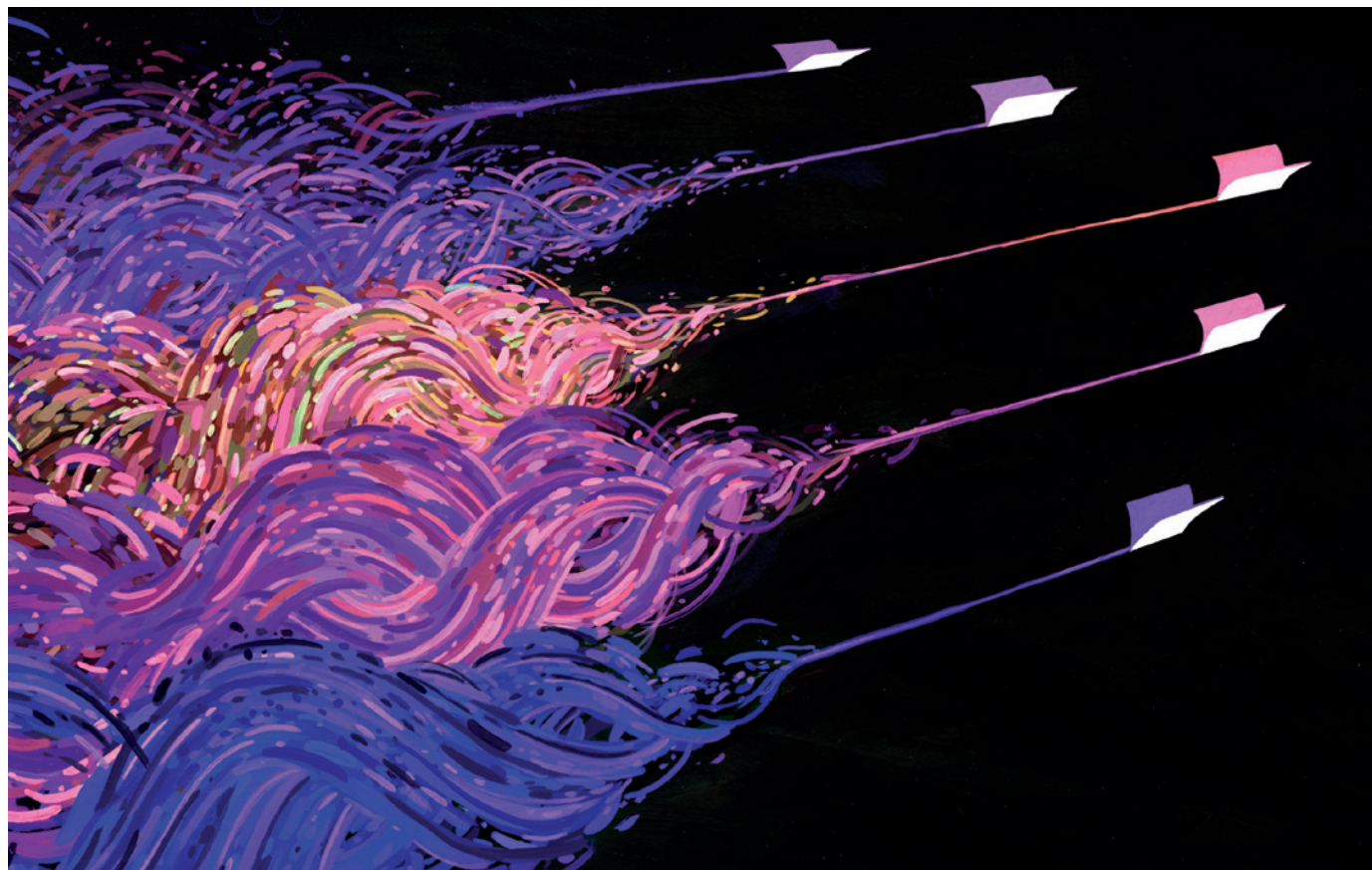


PUBLISHING Three advocates on how to hasten a free scholarly literature **p.442**

HISTORY The blighted communities at the heart of the nuclear arms race **p.444**

PUBLISHING A conversation about the new Digital Public Library of America **p.447**

BRENDAN MONROE



Beyond the paper

The journal and article are being superseded by algorithms that filter, rate and disseminate scholarship as it happens, argues **Jason Priem**.

Henry Oldenburg created the first scientific journal in 1665 with a simple goal: apply an emerging communication technology — the printing press — to improve the dissemination of scholarly knowledge. The journal was a vast improvement over the letter-writing system that it eventually replaced. But it had a cost: no longer could scientists read everything someone sent them; existing information filters became swamped.

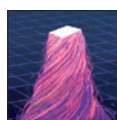
To solve this, peer and editorial review emerged as a filter, becoming increasingly standardized in the science boom after the

Second World War. This peer-review system applies community evaluation of scholarly products by proxy: editorial boards, editors and peer reviewers are nominated to enact representative judgements on behalf of their communities.

Now we are witnessing the transition to yet another scholarly communication system — one that will harness the technology

of the Web to vastly improve dissemination. What the journal did for a single, formal product (the article), the Web is doing for the entire breadth of scholarly output. The article was an attempt to freeze and mount some part of the scholarly process for display. The Web opens the workshop windows to disseminate scholarship as it happens, erasing the artificial distinction between process and product.

Over the next ten years, the view through these open windows will inform powerful, online filters; these will distil communities' impact judgements ▶



THE FUTURE OF PUBLISHING
A *Nature* special issue.
nature.com/scipublishing

► algorithmically, replacing the peer-review and journal systems.

THE PRESENT

This move towards a more diverse set of outputs has already begun. Scholars now share their research data in repositories such as GenBank, Dryad and figshare (figshare is supported by Digital Science, which is owned by the same parent company as *Nature*). They use repositories such as GitHub to share code, analyses and even ‘executable papers’ that automatically knit together preloaded data, analysis and prose. They challenge the traditional article format by including blog posts, interactive graphics and video. And perhaps most significantly, academics are moving informal scholarly conversations from the faculty lounge to social media platforms such as Twitter¹. In the next ten years, most scholars will join such networks, driven by both the value of improved networking and the fear of being left out of important conversations.

This shift from a paper-native to a Web-native system has three significant consequences. First, the flow of scholarly information is expanding by orders of magnitude, swamping our paper-based filtering system. Like journal articles, Web-native products are of inconsistent quality and require winnowing; unlike journal articles, the scale of these products overwhelms attempts at manual curation. In the late 1990s, commercial Internet services such as Yahoo! found that hiring experts to create vetted lists of web pages completely failed at Web scale; the same will be true of scholarship.

Second, the Web era is exposing the delicate tracework of ideas beneath the formal structures of the academy². The ephemera of scholarship are reified in a constellation of data points: views on figshare, mentions in discussions between colleagues on Twitter, saves in a reference manager such as Zotero or Mendeley, citations in an open-access preprint, recommendations on Faculty of 1000, and many more. We can use these and other tracers of impact to generate new metrics of scholarly influence. Informed by these alternative metrics, or ‘altmetrics’, we will draw new maps of scholarly contribution, unprecedented in subtlety, texture and detail. Suddenly, the rocky plain of ideas — once navigated using cairns of citation — is covered in fresh snow. In the Web era, scholarship leaves footprints.

The third consequence proceeds from the first two. The editors and reviewers employed as proxy community assessors will be replaced by the aggregated, collective judgements of communities themselves. The information-overload problem supplies its own solution. This is the power behind the Web’s great filter, Google. While Yahoo! was applying paper-native expert curation to the Web, Google was aggregating the collective

authority judgements inherent in the structure of the network. Its PageRank algorithm weights hyperlinks from authoritative sources more heavily. To find which sources count as authoritative, the same algorithm is applied to each of the source’s inbound links, and so on. This simple recursive algorithm has proved remarkably effective, and requires minimal manual tuning. It simply harnesses the quality judgements already being made by the community, implicit in their decisions to link to other pages. This core approach is also the future for scholarly communication.

THE FUTURE

Today’s publication silos will be replaced by a set of decentralized, interoperable services that are built on a core infrastructure of open data and evolving standards — like the Web itself (see ‘Reconstructing publishing’). This ‘decoupled journal’^{3,4} publishes promiscuously, then subjects products to rigorous review through the aggregated judgements of expert communities, supporting both rapid, fine-grained filtering and consistent, meaningful evaluation.

Dissemination. The concept of ‘pre-publication’ will be forgotten as frictionless, Web-based dissemination increasingly pervades the research process. Conversations, data collection, analysis and description will be born published. Many researchers are already practising this ‘open-notebook science’. For

“Tools are emerging to facilitate this ‘share early, share often’ approach.”

instance, population biologist Carl Boettiger describes his day’s research and provides in-progress code, analysis and writing, and chemist Jean-Claude Bradley’s lab at Drexel University in Philadelphia, Pennsylvania, publishes its entire output in near-real time. New tools are emerging to facilitate this ‘share early, share often’ approach. The journal *Push*, for example, lets scholars build journal articles incrementally, with each version tracked and open online, available for collaboration and comment throughout (see <http://push.cwcon.org>).

As the former constituents of the article — data, tables, figures, reference lists and so on — fracture and dissolve into the fluid Web, they will leave behind the core of the article: the story. Authors will create these stories like blog posts: lean, fast and heavily reliant on the published tools, data and analyses of the entire community. These stories will carry forward the synthesis and narrative functions of the traditional article. Like all decoupled journal products, they will be openly available, redundantly archived, easily attributable and enveloped in an open network of links, comments, citations, annotations, discussions, saves and other interactions.

Certification. Authors will select from among diverse paths of certification for their research products. For specialized products such as daily lab results, no certification is needed — use of these products will be limited to narrow subfields in which researchers are qualified and motivated to consume unfiltered results⁵. For products aimed at broader audiences, authors will rely on algorithmic filters to flag their work to others. *The Journal of Digital Humanities*, for example, does not take submissions; rather, it highlights the best content already published online, often pulling from relatively obscure blogs and web pages. Importantly, this selection process relies heavily on altmetrics (such as number of page views, tweets and trackbacks) as a first-pass filter before manual curation.

Similarly, forums such as MathOverflow use community-bestowed votes to promote or demote content. Many MathOverflow posts have been cited in the peer-reviewed mathematics literature, and some mathematicians have begun to report high MathOverflow reputation scores as evidence of scholarly standing.

Qualitative peer review will move into the open and become yet another public scholarly product — complete with its own altmetrics, citations and even other peer reviews. The reviewer will metamorphose from gatekeeper to interlocutor and collaborator.

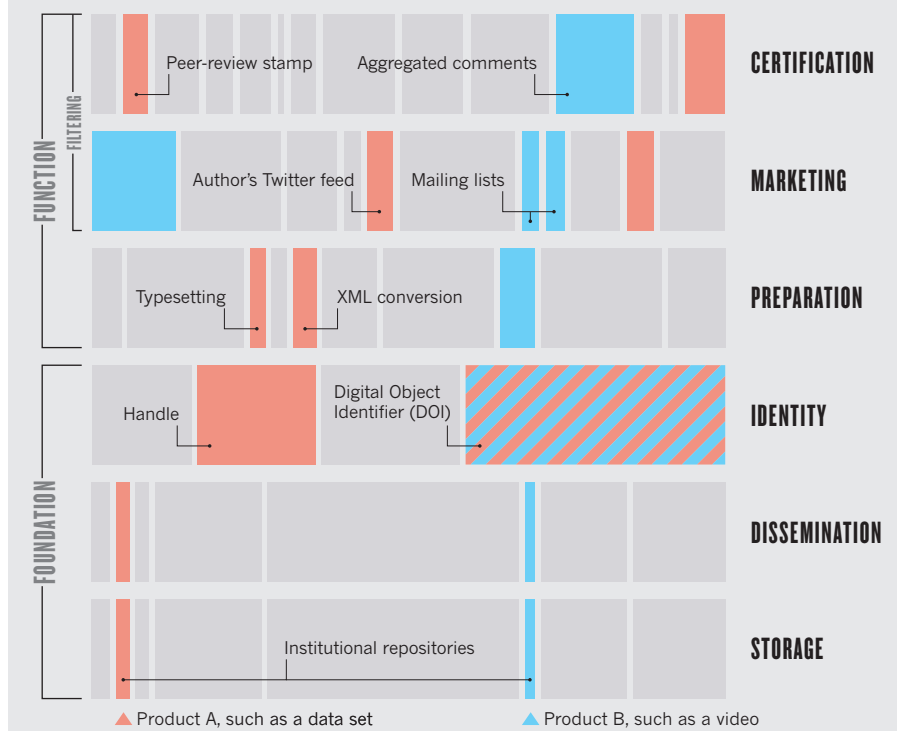
Start-up services such as Rubriq already offer ‘standalone peer review’, sold to authors as a tool to improve their work and help it to find its best audience. As peer review is unhitched from the creaky machinery of paper-native publishing, many more start-ups will offer other innovative variations. Some may specialize in statistical review or copy editing, or emphasize a quick turnaround. Many will offer automated review derived from altmetrics, while others will leverage existing journal brands.

Authors will select the best services for each product and purpose, pinning certifications to their products’ web pages like county-fair ribbons. For instance, an article might receive a ‘B+ overall significance’ badge from *Nature*, a ‘statistically rigorous’ badge from a specialist statistical review service and a ‘highly read’ or ‘highly influential’ badge from an altmetrics aggregator.

Filtering. To channel this torrent of Web-native scholarship, readers will rely on personalized recommendation engines to aggregate both altmetrics and qualitative badges, producing a bespoke, curated stream. Imagine every morning getting an e-mail from your personal filtering service with the five most important things for you to read that day. Content from your narrowest subfield — say, the evolution of toxins in *Scorpaeniformes* (an order of ray-finned fish) — is relatively unfiltered, resembling a simple RSS feed. At this level of detail,

RECONSTRUCTING PUBLISHING

In future, authors and funders will select services that are best for a given research product (for example, A and B). The product will be stored in and disseminated from an institutional repository, and given a unique identifier. Most traditional journal functions will be performed independently by an ecosystem of service providers. Some examples of services are labelled below.



information scarcity is as much a problem as information overload, and you want to stay aware of nearly everything your closest collaborators and competitors do. As you move out in scope (to venomous fish, then fish, then marine biology, for example), the content becomes increasingly filtered: you see only products that have been read, discussed, saved, cited and recommended by the other narrow subdisciplinary communities that are (like you) reading everything in their specific areas. These recommendations constantly improve as the system learns from your interactions (as Google search does already), but also allow for serendipity by delivering a customizable percentage of unexpected content from outside your narrow circles. Given such a system, why would you ever bother reading another journal?

Reward. The reward structure of scholarship will change too. Built on tools such as ImpactStory (of which I am a co-founder), scholars' CVs will transform from static lists of formal products to real-time displays of transparent, normalized metrics that reveal diverse impacts. Funders, ever eager to maximize value for their money, will demand more Web-native scholarly production. The US National Science Foundation, for example, has already begun requesting "products" rather than "articles" in the biographical-sketch section of grant applications⁶. The widespread availability of open altmetrics

will fuel this trend, as progressive scholars begin to tell — and to be rewarded for — compelling, data-driven stories about the impact of their diverse products.

Tenure and hiring committees will adapt, too, with growing urgency. Ultimately, science evaluation will become something that is done scientifically, exchanging arbitrary, biased, personal opinions for meaningful distillations of entire communities' assessments. We can start to imagine the academic department as a sports team, full of complementary positions (theorists, methodologists, educators, public communicators, grant writers and so on). Coming years will see evaluators playing an academic version of *Moneyball* (the statistical approach to US baseball): instead of trying to field teams of identical superstars, we will leverage nuanced impact data to build teams of specialists who add up to more than the sum of their parts. We will wonder at how meaningful evaluation could have been accomplished in the slow, coarse, paper-native world of the journal.

The peer-reviewed journal is the product of some 350 years of selective breeding, a thoroughbred exquisitely tuned to wring the best possible performance from the paper-based world of its creation. But the Web has irrevocably changed our information environment — it is no longer the habitat the journal evolved in.

The next decade will see a dramatic dieback

in journals and a Cambrian explosion of communication species evolved for the Web- and network-centric world. Most of these will be failures. We see this already in the dozens of 'online journal clubs' and 'Facebooks for science'; many are virtual ghost towns. But critics of these early failed systems miss the point. In the early twentieth century, there were more than 100 car makers; by mid-century, there were barely a dozen. But they were the dozen who had got it right. Similarly, we can expect the intense selection pressure generated by scholars' hunger for relevant information to drive rapid evolution in systems for certification, recommendation and assessment.

CONCERNS

The new era of scholarly communication will pose challenges. One significant concern is the ease of 'gaming' the algorithms we will rely on for filtering. What is to keep me from recruiting a few friends or students to tweet, read and save my new article, for example? Several companies have already emerged to sell social media follows and likes. Will I need to spend grant money on this sort of payola to compete?

Of course, any important metric will be gamed, just as citations are⁷. But altmetrics are not the easy target they may seem, thanks to their diversity and volume. Gaming a single metric may be easy; tampering with dozens is not. Products with high numbers of tweets but no downloads, for instance, will stand out. And not just to humans: given enough data, machines turn out to be excellent at finding discrepancies in standard usage patterns. Credit-card companies, for instance, use 'algorithmic forensics' to nab identity thieves. The powerful pattern detection of constantly evolving algorithms also maintains the robustness of Google search results and online advertising click counts — both of which are heavily gamed by highly motivated hackers⁸. These defences work in scholarly contexts, too: the Social Sciences Research Network and the Public Library of Science have already effectively used algorithmic shields around download counts⁹ (see go.nature.com/binjfk).

A second criticism is that the very idea of quantifying scientific impact is misguided. This really will not do. We scientists routinely search out numerical data to explain everything from subatomic physics to the appreciation of Mozart; we cannot then insist that our cogitations are uniquely exempt. The ultimate judge of scientific quality is the scientific community; its judgments are expressed in actions and these actions may be measured. The only thing to do is to find good measures to replace the slow, clumsy and misleading ones we rely on today. The great migration of scholarship to the Web promises to help us to do this.

A final concern is that in seeking the wisdom of crowds, we will end up governed

by the madness of mobs. Will a crowd-sourced scholarship be dominated by provocative pap that fills without nourishing?

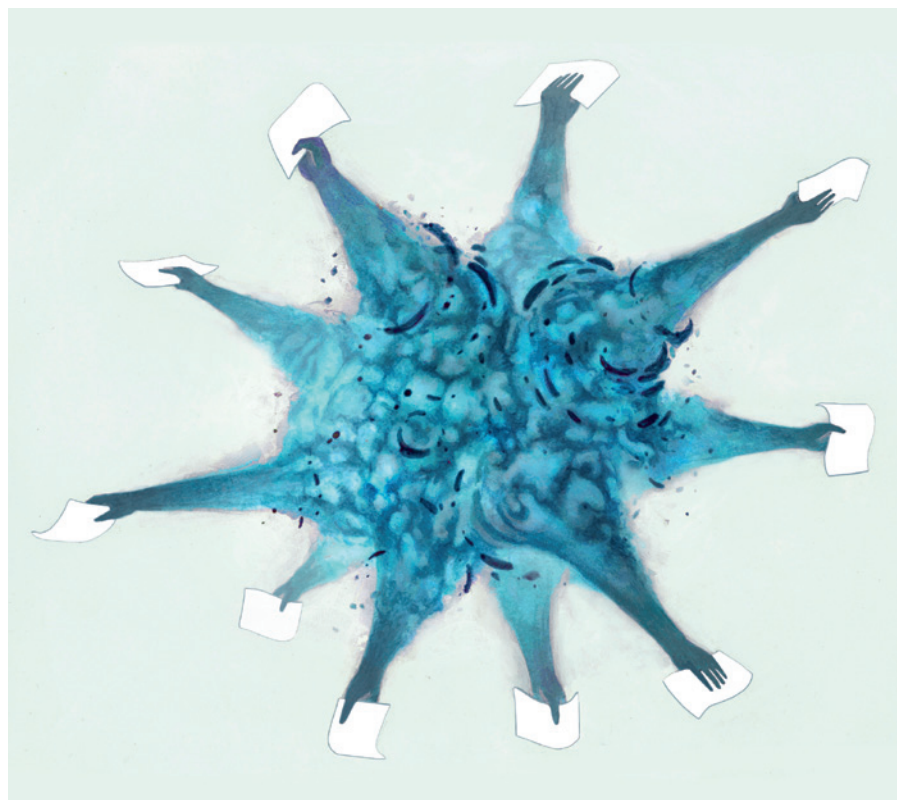
Here we must recall first that scholarship has always been a community enterprise, driven by building consensus among experts¹⁰. So the question is not 'Should we crowdsource?' but 'How should we crowdsource?'. Second, we must dispose of the straw-man argument that hundreds of uninformed readers' opinions will count for more than one Fields medallist's recommendation. Authority and expertise are central in the Web era as they were in the journal era. The difference is that whereas the paper-based system used subjective criteria to identify authoritative voices, the Web-based one assesses authority recursively from the entire community.

We now have a unique opportunity as scholars to guide the evolution of our tools in directions that honour our values and benefit our communities. Here's what to do. First, try new things: publish new kinds of products, share them in new places and brag about them using new metrics. Intellectual playfulness is a core scholarly virtue. Second, take risks (another scholarly virtue): publishing more papers may be safe, but scholars who establish early leadership in Web-native production will be ahead of the curve as these genres become dominant. Finally, resist the urge to cling to the trappings of scientific excellence rather than excellence itself. 'Publication' is just one mode of making public and one way of validating scholarly excellence. It is time to embrace the Web's power to disseminate and filter scholarship more broadly and meaningfully. Welcome to the next era of scholarly communication. ■

Jason Priem is an information scientist at the University of North Carolina at Chapel Hill, North Carolina, USA, and co-founder of ImpactStory.
e-mail: jason@impactstory.org

1. Priem, J., Costello, K. & Dzuba, T. *figshare* <http://doi.org/kvx> (2012).
2. Cronin, B. J. *Inform. Sci.* **27**, 1–7 (2001).
3. Priem, J. & Hemminger, B. H. *Front. Comput. Neurosci.* **6**, 19 (2012).
4. Smith, J. W. T. *Learn. Publ.* **12**, 79–91 (1999).
5. Esposito, J. J. *J. Electron. Publ.* <http://dx.doi.org/10.3998/3336451.0011.203> (2008).
6. Piwowar, H. *Nature* **493**, 159 (2013).
7. Wilhite, A. W. & Fong, E. A. *Science* **335**, 542–543 (2012).
8. Priem, J. & Hemminger, B. M. *First Monday* (5 July 2010).
9. Edelman, B. G. & Larkin, I. Harvard Business School NOM Unit Working Paper No. 09-096 (2009).
10. Kuhn, T. S. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1962).

The author declares competing financial interests: see go.nature.com/szgx9 for details.



BRENDAN MONROE

A fool's errand

Objections to the Creative Commons attribution licence are straw men raised by parties who want open access to be as closed as possible, warns **John Wilbanks**.

Copyright licensing is a topic usually left to law review articles, or obscure terms of service on websites, or agreements between publishers and libraries. But it is an essential element of the move towards open access — the free, immediate online availability of scholarly articles coupled with the right to use them fully in the digital environment.

An article that is free to read is not necessarily open for all uses — often, it cannot be reused for text mining or in derivative works, for example. The permitted uses depend on the copyright licence used by the author.

In my view, for an article to be considered truly open access, it has to meet the widely accepted definition in the Budapest Open Access Initiative — a set of recommendations laid out by leaders of the open-access movement in 2001. That is, users must be able to “read, download, copy, distribute,

print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.”

Traditional publishing licences tend to place restrictions on at least one of these uses, and it isn't easy for a reader to figure out what those are. If the reader is a computer, as is more and more prevalent, the restrictions are a spanner in the works.

The use of the Creative Commons attribution licence (CC-BY) fulfils the community definition of open access and avoids a future morass of articles with murky legal provenance and concomitant unclear reuse possibilities. CC-BY was launched in 2002, 2 years before I started a 7-year stint as head of science initiatives at Creative Commons in



THE FUTURE OF PUBLISHING
A Nature special issue.
nature.com/scipublishing

Mountain View, California.

CC-BY has now come under attack from the International Association of Scientific, Technical & Medical Publishers, which is discussing the introduction of a licence that would implement some — but not all — of the commonly accepted tenets of open access. At a conference run by the association in January, this was referred to variously as a “new” licence and even as “CC Plus”.

It is a bad idea. Here's why.

TRIED AND TRUE

CC-BY is a liberal licence that allows any kind of use under copyright as long as the author is credited in the manner in which he or she specifies. It is more than a decade old, clear, well-tested and deeply established as an effective open-access licence for both for-profit and non-profit publishers (see ‘Licence to share’). It has been translated into more than 50 languages and is legally enforceable around the world. No other open-access licence can claim its power, standing and adoption.

Critics have lately dubbed CC-BY ‘viral’, and bridled against the idea of research councils mandating its use as a way to implement open access for the scholarly literature. ‘Viral’ can be read either in the cultural sense — an article becoming wildly popular — or in the legal sense. The former is desirable. The latter is false: CC-BY does not force derivative works to be relicensed under the same terms.

Nor is it an unprecedented act for a funder to maximize its return on investment by specifying that publications arising from its funds be published under a liberal copyright licence. Taking money for research comes with conditions: grants from the US National Institutes of Health, for example, come with more than 70 such requirements, including data-sharing plans, annual reports, ethics training and gender indicators. That one of those conditions be intended to optimize a research article's impact by enabling its reuse by other researchers, their robots or by entrepreneurs, sits well within the funding tradition.

Any licence that is less open than CC-BY reserves the rights of the copyright holder to control certain reuses, and that requires a legitimate justification. If an article's publication costs have been covered by an article-processing fee, then reserving rights is just a means of double dipping.

Furthermore, funders who want the maximum impact are going to choose an existing standard legal tool that is interoperable with the vast majority of free culture and free software licences, not bet on a licence that might not be. Licences that distinguish between kinds of reuse, or discriminate against entrepreneurs, fail every definition of open access, open knowledge and open source. Tiny details of drafting, intentional or not, often render content under one licence legally incompatible with content under a different

one. An essential function of the limited set of Creative Commons licences is to forestall the hobbling impact that licence proliferation has on the network effect of open culture.

On my cynical days, I fear that this kind of hobbling is at the heart of a strategy to create ‘open-access’ licences just for scholarly publishers. These licences would reserve the most creative reuses for those who simply serve as the midwives for content, not for those who might go on to create works that can surprise, inform and delight. If we allow only a tiny set of predicted reuses, those are, by definition, the only reuses we will get — and they will benefit only the existing power players in scholarly publishing.

SPECIOUS CONCERNS

It is hard to precis all the specious concerns about why CC-BY will not work for scholarly publishing. In brief, those opposing the licence often claim that it: would allow others such as drug companies to sell works downstream; implies author endorsement of shady overlay journals; would require all the elements of an article to be freely licensed (including photographs, music, modern art and, presumably, Hollywood films); and would make attribution on text mining unwieldy.

CC-BY does indeed allow resale — of something that is already on the Internet for free. Anyone who pays for an object under CC-BY is either making a donation, or is paying a tax for being inept at searching the Internet. And a few key elements of CC-BY make it possible to prevent dastardly uses.

First, because attribution is up to the author and the journal, it is easy to make sure that any copy — if someone is trying to make a shady business of reprints, for instance — links back to the free version of the article.

Second, CC-BY does not grant publicity rights. That means that attribution can be used to clearly disclaim endorsement without

the tangles caused by commercial restrictions. The attribution can carry a requested form of citation, including the URL to the original, peer-reviewed, free and branded copy of the article. Then, anyone wishing to reprint must also reproduce that citation or be in violation of the copyright. It makes it hard to imply endorsement or sell something when the object itself carries the provenance, and links, to the free version in a trusted journal.

Similarly, publishers can communicate their desired attribution in a text-mining context. Indeed, because a text miner is only extracting ‘facts’ from the text, those facts are

“When in doubt, use running code that someone else has already written.”

by law not covered by copyright — and thus not subject to the attribution requirement. Of course, from a technical and scientific perspective, readers will always want to know the provenance of a fact, and it is good

practice to link back to the source. Indeed, the idea of link-based provenance is built into the design of the Semantic Web. It is a technical problem, not a legal one.

It is also easy to include in an article under a CC-BY licence items not subject to the licence — images, musical notation and more. That simply requires the rider “this article, unless otherwise noted, is available under CC-BY”, and a note placed by the elements that are not available. Thus, a photograph under a Creative Commons licence can be used as one of *Time* magazine's Photos of the Year, for example, without rendering the rest of the publication subject to the terms of the licence.

In sum, the debate over CC-BY is actually an attempt to reinterpret what the ‘open’ in ‘open access’ means. There are parties who want open access to be as closed as possible to protect their business models from change.

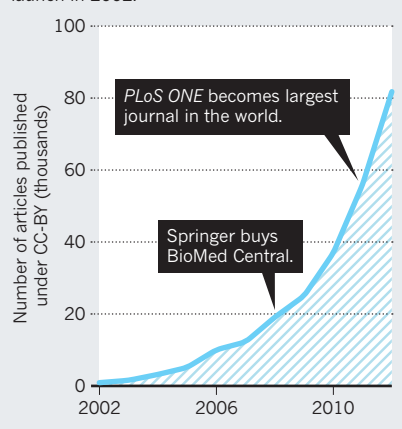
A licence that is designed just for publishers might feel safer, but it is a fool's errand. CC-BY exists. It is used widely and is driving tens of millions of dollars in annual revenue for scholarly publishers such as BioMed Central, Hindawi and the Public Library of Science. It fulfils the community definitions of openness. And it works with the vast web of existing free content. For any new licence to achieve all that, it would need to be CC-BY.

It is encouraging that the open-access debate has moved from ‘should we?’ to ‘how do we?’, and that we're talking about the issue even in this august publication. It would be deeply sad if we were to fail now to draw on one of the basic lessons of the open movements that have come before: when in doubt, use running code that someone else has already written. ■

John Wilbanks is the chief commons officer at Sage Bionetworks in Seattle, Washington. e-mail: john.wilbanks@sagebase.org

LICENCE TO SHARE

Leading open-access scholarly publishers have increasingly adopted CC-BY since its launch in 2002.





How to hasten open access

Three advocates for a universally free scholarly literature give their prescriptions for the movement's next push, from findability to translations.

ALMA SWAN

Time to align policy

Director of European advocacy, Scholarly Publishing and Academic Resources Coalition (SPARC) Europe

Policies for open-access publishing have made rapid progress, raising a pressing issue — if the research literature is to become completely open within the next 5 years, different policies must be aligned. For researchers who are funded by more than one source, this harmonization is crucial in creating a single, simple path to making their work open access, affordably and sustainably.

Funders, universities, nations and states need to work together now to secure good open-access policy. What does that look like? It is clear, accurate and focused. It supports new models for research communication. It does not let copyright become a hurdle and does not interfere with authors' choice of journals. It mandates immediate access. It does not perpetuate a broken market, but finds ways to curb costs.

Good policy doesn't refer to 'gold' open



access as 'author pays' or 'paid for', because 66% of open-access journals do not charge. It uses the proper definition: journals that make their content immediately and freely available on the Internet. It doesn't call 'green' open access 'embargoed access', because 60% of the time it is not. It defines it as literature that is made open access directly by the author, usually through a repository. It doesn't assume that green open access harms publishers, because evidence shows that it does not.

Good policy exploits the fact that gold and green routes can both provide articles with licences that maximize re-use (see page 440).

It introduces measures of compliance — and sanctions for non-compliance — because these shape behaviour. It embraces the fact that open access will cause disruptive change. And it puts the interests of research, and of the public that pays for it, before all others.

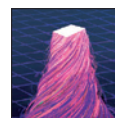
Specifically, funders and universities must work together to align policy with respect to permissible embargoes, timing of repository deposits for green open access and whether research grants can be used to buy gold open access.

They must consider whether payments can be made to 'hybrid' journals (subscription journals that make articles open access for a fee) and the position that funders take on rights, licensing requirements and which repositories can be used.

The European Commission (EC) has been explicit about its intentions on these details. From January 2014, it will require articles to be placed in institutional repositories immediately upon publication or acceptance. It permits only a six-month embargo in disciplines within science, technology, engineering and medicine, and a 12-month embargo in the humanities and social sciences. The policy encourages authors to retain copyright and to give publishers licences to publish their work, and it allows grant funds to be used for publishing charges. The EC has stated its hope that European countries will develop open-access policies in line with its own.

State-level policy proposals in the United States align with proposed federal policy in terms of permitted embargo times (the maximum is 12 months). In Belgium and Ireland, funder and university policies are aligning their focus on deposits in repositories, requiring this for research assessment and echoing the EC's maximum permitted embargo times. In Spain, Norway and Denmark, policies that align with that of the EC are already in place. The two Australian funders' policies are similar, calling for deposit in repositories and differing only slightly on permitted embargoes.

So, harmonization is happening. But the development of seamless policies among funders needs to be the focus of advocacy. It will mean less author confusion and greater compliance. Authors will begin to understand the potential of truly open research and be inspired to devise further innovative practices. Then we can expect true disruption: a very different kind of scholarly communication, catalysed by good policy.



THE FUTURE OF PUBLISHING

A Nature special issue.

nature.com/scipublishing



MATTHEW COCKERILL Make indexing fast and fair

Managing director, BioMed Central

Imagine a world in which Google demands that websites prove their worth for several years before their pages become searchable. In essence, that is the model that is currently in place for scientific publishing. For a journal to be indexed within PubMed Central, Web of Science and Scopus — in other words, for its publications to be easily visible to scientists — it must provide a track record of high-quality publications. The catch-22 is that while proving itself, that journal cannot offer authors one of their most important requirements: discoverability.

For open access to take flight, the indexing and visibility of new journals needs sorting out. PubMed Central's original policy of rapid inclusion of open-access journals was crucial in helping the model to prove itself. The retreat in 2009 from this inclusiveness has threatened to stymie evolution in science publishing, helping incumbents to defend their position and disadvantaging some new entrants.

The controversy generated when the journal *eLife* jumped the queue by appearing in PubMed Central before having published a single paper, let alone the current usual requisite of at least 15, highlights how baffling entry criteria have become. It would be a disservice to the community for a high-quality journal such as *eLife* not to be promptly tracked. But the jump-start gave the unhealthy impression that, to paraphrase George Orwell, 'all journals are treated equally, but some are treated more equally than others'.

Conversely, one of the leading journals we publish at BioMed Central, *Genome Medicine*, has only just been added to Web of Science by Thomson Reuters 4 years after its launch. The delay occurred despite a strong editorial board, a steady publication rate of around 90 articles a year and a predicted impact factor of around 6 — greater than those for 95% of the journals included in Web of Science.

'Predatory publishers' — those offering journals of low quality — are undoubtedly a concern (see page 433). But endlessly raising the walls of the citadel, or making those walls scalable only by journals with friends in high places, is not the answer: it slows scientific progress.

Instead, indexing services should err on the side of including content from any new journal that meets basic standards, perhaps flagging it with a 'provisional' tag, but ensuring that it is discoverable and citable. If journals are subsequently found to have poor editorial standards or excessively low citation rates, they can easily be dropped from the index, and prior content removed or flagged with a warning.



DOUGLAS SIPP Translate local journals

RIKEN Center for Developmental Biology, Kobe, Japan

Once open-access publishing becomes common practice, citizens of English-speaking countries will enjoy the unearned advantage of being able to read and use research funded by other nations. But those nations' populations may be unable to read articles reporting research funded by their own taxes.

Governments and public-funding agencies in such nations will soon find themselves pressed to resolve this inequity.

In the major east Asian scientific powers of China, Japan and South Korea — the world's 2nd, 4th and 12th most-productive nations in terms of numbers of research publications — research and funding institutions have established open repositories for published research. The National Science Library of the Chinese Academy of Sciences, for example, mandated in 2009 that its members file articles in its open repository within one month of publication. At present, this seems to be used mainly for articles in Chinese-

language journals. Japan's largest funder of basic scientific research, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), proposed only last year that researchers receiving MEXT grants should report the access status of their publications. It has not specified any access requirements.

Japan and South Korea have both invested in science communication and public-understanding efforts to improve their citizens' access to scientific information, through better funded science museums, media and online programmes and training of communication and outreach specialists. For example, in 2001 Japan opened a large science museum, the Miraikan, and there are now programmes in science communication and journalism at Hokkaido and Waseda universities. South Korea has increased funding for its national science museum, and the Korean Academy for Science and Technology in Seongnam now offers media training.

But these activities fall short of unmediated access to full sets of data, methods and analyses. Local-language journals, another partial solution, remain uneven in quality and low in impact. English is still undisputedly the common language of the sciences.

The clearest way to achieve true open access in non-English speaking nations might be to borrow models from the open-access movement itself. A green open-access approach to lowering the language barrier could see funders call for scientists or institutions to produce local language translations of any published work conducted using public funds and host them in open repositories. A gold open-access approach, by contrast, might subsidize translations by journal publishers or other third parties. In either case, the solution will be costly and time-consuming, and require countries where English is not the native language to work harder than their anglophone counterparts to ensure public access to one of humanity's most precious resources — its knowledge. ■

M.C. has declared competing financial interests: see go.nature.com/ns5f9q for details.



Plutonium samples trundled around Hanford, the world's first plutonium plant, in 1954.

MILITARY HISTORY

Dinner at the Fission Chips

Mark Peplow assesses a chronicle of the blighted US and Soviet communities that fuelled the nuclear arms race.

“We had only washcloths, buckets, and sometimes rubber gloves,” recalls Faina Kuznetsova, a lab technician who routinely dealt with major leaks at Russia's first plutonium-manufacturing plant. “We mopped up the spills and poured them into big glass bottles. It was a very expensive compound and we were expected to recover every drop.”

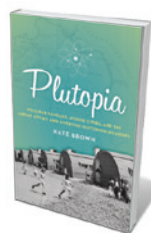
A multitude of such harrowing accounts fills the pages of *Plutopia*, a ‘hidden history’ of two communities — one American, one Soviet — that fuelled the nuclear arms race. Unusually, historian Kate Brown interviewed dozens of frontline workers for her meticulously researched account of how these two remote towns became indelibly linked by plutonium, and by catastrophic radioactive contamination.

Brown's story begins in December 1942, when the Manhattan Project to build nuclear weapons was in full swing, and the crucial ingredients — enriched uranium and plutonium — were in short supply. The US Army selected Hanford in Washington

state as the site of the world's first plutonium plant: it had plentiful water and hydroelectric power from the Columbia River, and the site covered 160 square kilometres of bare drylands.

The plant's first reactor fired up in autumn 1944, blasting slugs of uranium-238 with neutrons to create plutonium-239. Less than a year later, on 9 August 1945, the Hanford plutonium found its target via ‘Fat Man’, the atomic bomb that destroyed Nagasaki three days after the less-advanced uranium fission bomb ‘Little Boy’ detonated over Hiroshima.

As weapons stockpiles grew, Hanford's growing workforce was housed in Richland,



Plutopia: Nuclear Families, Atomic Cities, and the Great Soviet and American Plutonium Disasters
KATE BROWN
Oxford University Press: 2013. 416 pp.
\$27.95, £17.46

one of Brown's ‘plutopias’: a stage-managed American idyll where everyone was white, middle class and loved the bomb. Yet Richland's chintzy Americana — the Fission Chips restaurant or the high school with the mushroom-cloud mascot — concealed a calculated programme to keep residents ignorant of the risks.

These were considerable. Hanford's operators used the surrounding countryside as a vast dumping ground. They reasoned that dissipating the radioisotopes in river water, or the atmosphere, was the best way to dilute any danger. Radioactive waste poured daily into the environment, while hundreds of millions of ‘hot particles’ of radioactive metal fell like deadly snow across Richland. Meanwhile, the town grew into a corporate dictatorship, where plant managers controlled citizens' lives in a way that would have made a Soviet general proud.

Time and again, official review committees denied the evidence. In the early years, site medics tested workers after long weekends, when most of the radioisotope contamination had already been urinated from the body. In 1958, the Atomic Energy Commission reported to Congress that from 1944 to 1958 there had been only one radiation-related injury at Hanford, out of 18,000 workers. In fact, by 1959, infant deaths in Richland were four times the state average.

A pivotal point was the physicist Ernest Sternglass's 1963 paper in *Science* showing a link between maternal exposure to X-rays and child mortality from cancer, with implications of similar effects for radioactive fallout. Independent studies followed, and by 1966 the radioactivity in Richlanders' bones had been found to be 50% higher than normal.

Meanwhile, the Soviet leaders had taken note, and in 1945 began building their own plutopia in the remote Urals. In what Brown describes as the “Bronze Age beginnings of the Soviet atom”, conscripted labourers endured terrible conditions to build the Mayak plutonium plant and the closed city of Ozersk. The first reactor produced enough plutonium for a bomb by 1949, but many of the fuel slugs, loaded with the Soviet Union's entire stock of uranium, had cracked. So Mayak's bosses ordered workers to unload and sort 39,000 irradiated slugs by hand, protected only by a fortifying glass of vodka.

Thousands of exposed workers routinely suffered nausea, nosebleeds and intense pain, followed by crumbling bones, cancers and death, Brown reveals. To quell the growing unrest, Ozersk was showered with cinemas and apartments filled with modern conveniences, and then touted as proof that Communism delivered prosperity — just as

federal dollars were shaping Richland into the very model of the American dream.

Soviet plant managers also copied, and surpassed, the Americans' approach to waste disposal. By 1951, 20% of the nearby Techa river was radioactive effluent from the plant, flowing through dozens of towns in which more than 124,000 people lived. Then, in 1957, an underground waste-storage tank exploded, belching forth a mushroom cloud that irradiated hundreds of thousands of people.

Ozersk's residents cleaned up the plant with wire brushes and hoses — almost certainly receiving severe radiation doses. There is no agreement on how many died as a result of their exposure. Today, the region is a vast radioactive swamp — yet Mayak continues to process radioisotopes, and managers plan to keep dumping radioactive waste into open reservoirs until at least 2018, Brown says.

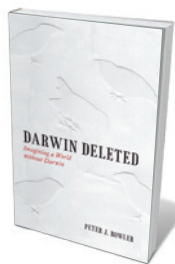
Back in Richland, whistle-blowers and investigative journalists picking away at the town's anodyne facade forced the US government to reveal the true scale of Hanford's contamination in 1986. Just weeks later, the catastrophic explosion at the Chernobyl nuclear power plant in Ukraine occurred. In the face of this double whammy, Hanford was slated to close. Together with the Mayak plant, it had released much more radioactivity into the environment than Chernobyl and the site will take half a century to clean up. Yet most Richland residents fought hard against Hanford's critics. They loved their town's social homogeneity, orderly management and relative prosperity, and feared that anti-nuke whiners would cost them their apple-pie lifestyle.

The only underwritten character in *Plutopia* is plutonium itself. Brown never explains why bomb-makers preferred plutonium over uranium (it is easier to purify, and requires a smaller critical mass), how it was processed or how fission works. And although radiation physics is swamped with confusing and archaic units, Brown fails to guide the reader through rems, rads, roentgens and curies, when a little context could help to make sense of the numbers.

Nevertheless, *Plutopia* has important messages for those managing today's nuclear facilities, arguing for caution and transparency. Highly subsidized communities are still a feature of civilian nuclear programmes in many countries, including Japan. It is no coincidence, Brown suggests, that the Fukushima nuclear accident in 2011 was characterized by poor safety protocols, official denials and a heavy use of underpaid workers to clean up the mess. ■

Mark Peplow is a science journalist based in Cambridge, UK.
e-mail: peplowscience@gmail.com

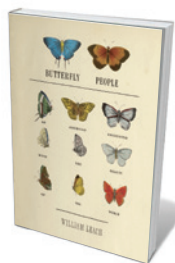
Books in brief



Darwin Deleted: Imagining a World Without Darwin

Peter J. Bowler UNIV. CHICAGO PRESS 336 pp. \$30 (2013)

Is there anything to add on Darwin and his work? In his 'counterfactual' history, Peter Bowler manages it neatly, wondering what might have happened had Darwin not published his theory of evolution by natural selection. Bowler argues that at the time, only Darwin had the originality and largeness of vision to craft his big idea, but its very boldness polarized thinking. The theory, Bowler surmises, would have emerged *sans* Darwin, but later — which, ironically, might have eased the broad acceptance of evolution.



Butterfly People: An American Encounter With the Beauty of the World

William R. Leach PANTHEON 416 pp. \$32.50 (2013)

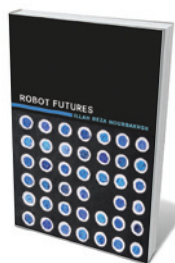
Butterflies in their thousands blanketed summer fields in nineteenth-century America. These 'flying jewels' drew a generation of amateur natural historians from the cultural chrysalis. Here, historian William Leach celebrates several — including Herman Strecker and Samuel Scudder — who created a home-grown field. Fed by Linnaeus, Darwin and crowd-sourced photographs and specimens, this scientific search for beauty collided early on, however, with the country's commercial drive.



Picking Up: On the Streets and Behind the Trucks With the Sanitation Workers of New York City

Robin Nagle FARRAR, STRAUS AND GIROUX 304 pp. \$28 (2013)

Daily in New York city, around 9,000 people clear away 11,000 tonnes of household waste. In her 10-year, sometime-firsthand study of 'san man' crews, cultural anthropologist Robin Nagle shines a light on their invisible lives. She reveals them as agents of urban reform and public health; traces the history of sanitation in the city, starting with eighteenth-century reformer Cadwallader Colden's yellow-fever control; and evokes the physical and psychological toll of this dangerous, filthy, necessary work.



Robot Futures

Illah Reza Nourbakhsh THE MIT PRESS 160 pp. \$24.95 (2013)

This glimpse into the future of robotics hums with enthusiasm. In his work, roboticist Illah Reza Nourbakhsh has created a raft of objects and capabilities, from robot three-dimensional visioning systems to a highly propulsive pogo stick. Here, prefacing each chapter with an imagined scenario, he forecasts how bots will invade commerce, the home and the human body. The possibilities — such as therapeutic, injectable robot colonies — are often provocative, but tempered by astute insights into the ethical and social implications of a roboticized world.



Narwhals: Arctic Whales in a Melting World

Todd McLeish UNIV. WASHINGTON PRESS 216 pp. \$26.95 (2013)

The tusked, deep-diving, upside-down-swimming narwhal is a cetaceous enigma. Questions hang over its feeding habits, strange dental arrangements, population and migration. In this portrait of the species, Todd McLeish mixes research, observations from High Arctic trips and engaging detours into iceberg ecology and more. While the debates rage on — over the tusk as a sensory organ, for instance — many agree that the animal's adaptive capacity may not keep pace with the shrinkage of sea ice.

Down the hatch

David Katz savours a guided tour of the ins and outs of the gastrointestinal tract.

A singular recollection came to me as I began the adventure of *Gulp*. I used to teach a course called 'Clinical Concepts in Public Health' at the Yale School of Public Health in New Haven, Connecticut. This was my opportunity to convey to bright young graduate students the marvels of the human body in both function and dysfunction. It was an epiphany. Compressing the greatest hits of medical education into a single semester helped me to fully appreciate how truly astounding it all was.

My impression is that Mary Roach experienced her guided tour of the gastrointestinal tract in much the same way. There is a real sense that her mind boggled at its wonders. In her blend of excellent investigative journalism, history, popular science and stand-up comedy, you are in for diverting surprises: don't expect this book to be a straight shot from the northern to the southern orifice. Each port of call along the way is more an excuse for fascinating detours than a destination in its own right.

We learn, for instance, about the work of some of the world's more specialized and surprising scientists, past and present. One has devoted a career to studying what causes variance in the odours of flatulence; another researches the secrets of saliva, such as what makes it flow. We learn about a scientist who pushed a live eel through a fistula — a surgically created hole — in a dog's stomach, so that only the eel's head protruded, to determine the extent to which animation upsets the actions of digestive juices. The answer proved to be: not at all. The eel was digested; the scientist, divorced.

We learn how snakes digest prey they have swallowed whole (through extremely powerful enzymes) and the likely cause of Elvis Presley's death (Hirschsprung's disease, or toxic megacolon). Roach reveals why the various worm species that are swallowed alive by reptiles and amphibians cannot eat their way back out. We are told the possible origins of the myth of fire-breathing dragons (hint: it has to do with decomposing mammals in the guts of large snakes, in proximity to Stone Age campfires).

In *Gulp*, Roach imparts her myriad insights eloquently, and with acerbic wit. A dog's nose

"takes in the sights by smell", and "it takes approximately as long to chew narwhal as it does to hunt them". She regales us with prime examples of nominative determinism: a scientist named Spitz studies saliva; one called Grime researches laundry detergents.

Adroitly avoiding squeamishness, Roach reveals the coda to this story — the anus — as a remarkable orifice. She shows how sensitive it has to be to differentiate solid, liquid and gas, and to perform its yeoman's work accordingly. As Robert Rosenbluth, a medic she quotes, puts it: "No engineer could design something as multifunctional and fine-tuned as an anus. To call someone an asshole is really bragging him up." (Personally, I find that statement smacks of a slight



Gulp: Adventures on the Alimentary Canal

MARY ROACH
W. W. Norton/
Oneworld: 2013.
352 pp. \$26.95,
£11.99

excess of enthusiasm, and take umbrage on behalf of my kidneys and liver, to say nothing of my brain.)

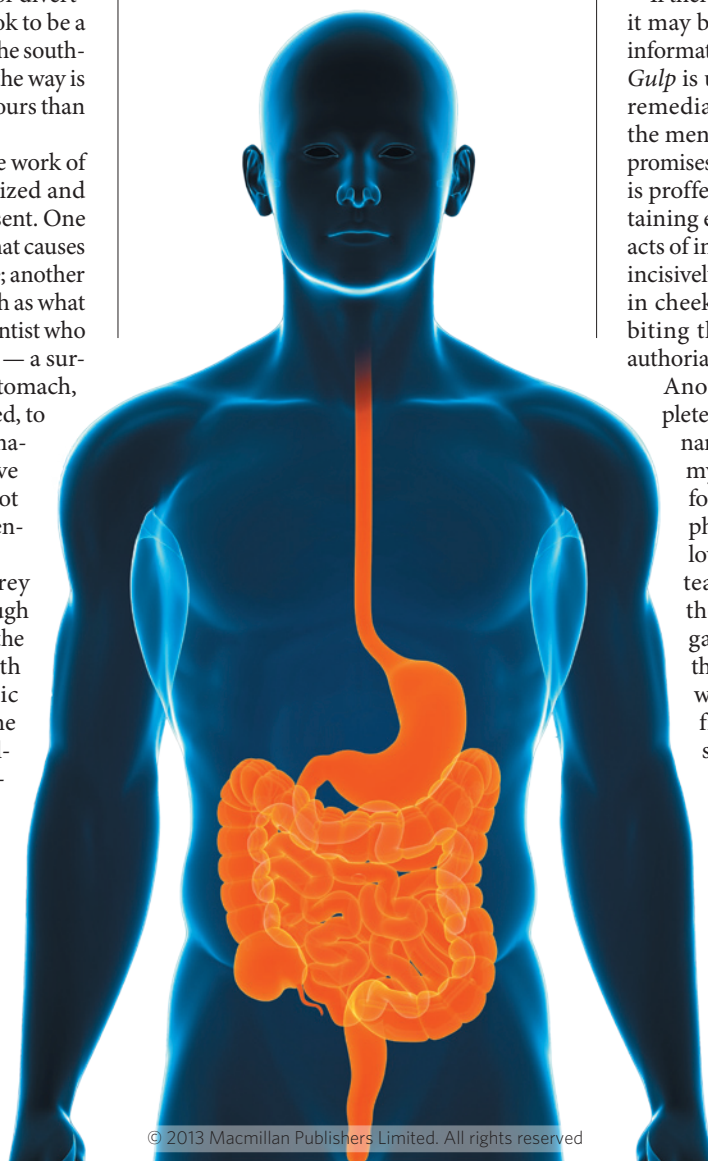
The last chapter addresses the gut's role in autoimmunity, and discusses the potential therapeutic value of 'fecal transplants' in treating colitis caused by the bacterium *Clostridium difficile*. A promising approach to a difficult problem, Roach tells us, assum-

ing we can get past the 'ick factor'.

If there is anything at all unpalatable here, it may be the relative paucity of practical information. If something ails your innards, *Gulp* is unlikely to help. There is nothing remedial, therapeutic or prescriptive on the menu here — but Roach has made no promises to offer them. Her food for thought is proffered purely in the service of entertaining enlightenment. Roach lays out the acts of ingestion, digestion and egestion, all incisively chewed over — with tongue ever in cheek. How she does so without once biting that tongue is a testament to her authorial gifts.

Another memory surfaced as I completed the journey along this extraordinary canal. Some 20 years ago, during my medical residency, I wrote a skit for an annual roast of our attending physicians. There were several well-loved gastroenterologists among the teaching faculty, and I lampooned their avid enthusiasm for all things gastrointestinal by asserting that they were, on the side, producing the world's first science-fiction movie filmed entirely through a colonoscope: *There's Life on Uranus!* It's a film Roach might have directed. ■

David Katz is the founding director of Yale University Prevention Research Center in Derby, Connecticut, USA, editor-in-chief of the journal *Childhood Obesity* and president-elect of the American College of Lifestyle Medicine.
e-mail: david.katz@yale.edu





BRYCE VICKMARK/REDUX/EYEVINE

Q&A Robert Darnton

Knowledge liberator

Robert Darnton heads the world's largest collection of academic publications, the Harvard University Library system. He is also a driver behind the new Digital Public Library of America. Ahead of its launch in April, he talks about Google, science journals and the open-access debate.

How did you come to run the Harvard libraries?

As a graduate student at Oxford in 1963, I began writing about books in revolutionary France, helping to found the discipline of book history. I was in my academic corner writing about Enlightenment ideals when the Internet exploded the world of academic communication in the 1990s. As president of the American Historical Association I started a programme to make dissertations into e-books in 1999. Before I knew it I was involved in other electronic projects. Harvard invited me to become director of the libraries in 2007. From that position I have been trying to help grow the digital landscape to serve the public good.

Did you get involved in Google's effort to scan books?

In 2002 Google began an ambitious project to digitize every book in the world. It was intended as a search project: type in a query, and Google would show you snippets. They asked university libraries for books, which they would scan for free. At Harvard we didn't permit them to take works under copyright, but other libraries gave them everything.

What happened then?

In 2005 Google was faced with a lawsuit from authors and publishers claiming copyright infringement. After three years of negotiations, they came back with a settlement for a

commercial library. Google would split the profits with copyright owners, leaving libraries to buy back digital subscriptions to their own books. In 2011, a federal court declared the settlement unacceptable. Now Google has a great digital database of books, which it can use in cooperation with publishers to sell books, but in my view it's not creating the kind of library that was promised.

Digital Public Library of America launch
Boston Public Library,
Boston, Massachusetts.
18–19 April 2013.

So you're launching a public alternative?

Along with research libraries, archives and museums throughout the country, Harvard is supporting the non-profit Digital Public Library of America (DPLA). At its launch in April, the DPLA will make freely available online large runs of documents from digitized special collections. Harvard will contribute, among many previously inaccessible documents, 243 rare medieval manuscripts and 3,628 daguerreotypes, including the first photographs of the Moon. This initial core of offerings will probably grow to include about 5 million books in the public domain and perhaps 2 million 'orphan' books whose owners cannot be found. We have also made

our data compatible with Europeana, a digital library that gathers collections from 27 countries.

Will the DPLA contain science?

We do not currently have plans to digitize textbooks or journals. But sometimes the raw data are more important: scientists no longer leave a paper trail. Digital data are more fragile than printed material. Eventually we want our online library to preserve raw experimental data, not just published results.

Why not science journals?

In recent years their prices have risen four times faster than inflation. An average chemistry journal subscription is more than US\$4,000 annually. An institutional subscription to the *Journal of Comparative Neurology* costs more than \$29,000 per year. Libraries used to spend half their budget on journals; now many spend 90% and can't afford monographs.

How did this situation arise?

Academic publishing grew up in the nineteenth century, when professional associations needed to disseminate their members' works. Recently, publishers have taken over, creating thousands of journals so specialized that there is little competition. Most research is funded by taxes, but the taxpayers do not have free access to the results of the work that they financed. Academics do the research, write the papers and referee them — all for free — then buy back their work at exorbitant prices. Publishers can rake in the money, because libraries can't unsubscribe without alienating scientists. It wasn't a conspiracy or the result of wicked aims. But the whole system is fundamentally flawed and we must do something about it.

What is Harvard doing to fix it?

The Harvard faculty has decided that the price increases are unsustainable. We have 73 libraries and each must limit the haemorrhage of funds. In April 2012, a faculty council asked faculty and students to consider not contributing to journals that charge outrageous prices. But voluntary effort alone won't solve the problem. We've created a repository where all Harvard professors by default must deposit their articles, although they can opt out, and the traffic from around the world is astonishing. The university will pay most of the processing fees for articles submitted to open-access journals such as those of the Public Library of Science. By working with other universities, we hope to make most journals open-access and eventually to change the economics of journal publishing. Open-access journals get more hits than closed ones. This is where the future lies. ■

INTERVIEW BY JASCHA HOFFMAN



THE FUTURE OF PUBLISHING
A Nature special issue.
nature.com/scipublishing

Correspondence

Online integrity training falls short

Education in the responsible conduct of research (RCR) will receive a long-overdue critique at a conference discussing the work of the US Office of Research Integrity on 3–5 April in Baltimore, Maryland.

The US National Institutes of Health has required recipients of training grants to receive RCR education since 1990, and it has been a prerequisite of the US National Science Foundation (NSF) since 2010 for all students and postdocs funded by its research grants.

In 2012, under contract from the National Center for Professional and Research Ethics at the University of Illinois, I reviewed the NSF policies of 27 major universities. I found that 26 depend solely (12) or largely (14) on online RCR training, with all but two using the Collaborative Institutional Training Initiative (CITI; www.citiprogram.org).

Outsourcing ethics education in this way suggests that RCR education is developed and executed with an eye to expedience rather than excellence. It risks sending a message to young researchers that the university and its scientists do not place much emphasis on responsible conduct.

European universities, which do not yet have RCR mandates in place (see N. Axelsen and X. Bosch *Nature* **489**, 208; 2012), should learn from the US experience and develop meaningful RCR programmes. These need to be taught by the people the students want to emulate — scientists.

Kenneth D. Pimple *Indiana University, Bloomington, USA.*
pimple@indiana.edu

Too much reliance on anonymous tip-offs

Several scientific journals and ethics committees are deferring to anonymous judgment when

it comes to charges of plagiarism and falsification of results in published research papers. As a bioethicist, I believe that this practice is risky, even when the tipster's views are valid: it could itself damage the integrity of scientific research.

The authors of the blog Retraction Watch (www.retractionwatch.com) hold the view that anomalies detected by someone using the pseudonym 'Clare Francis' are useful to scientific journals, irrespective of his or her anonymity (see A. Marcus and I. Oransky *Lab Times* **7**, 39; 2011). But this unorthodox 'review' process pollutes the ethics that underpin scientific progress.

The practice could stimulate witch-hunting and pillorying. There is a danger that research-integrity committees could succumb to moralistic drift and confuse errors with misdeeds, underestimating context and a scientist's professional record. Stigmatized researchers might be tempted to exact revenge on their colleagues.

In my opinion, this is not the way to improve the moral standards of science's contribution to society or to build public engagement in science.
Gilberto Corbellini *Sapienza University of Rome, Italy.*
gilberto.corbellini@uniroma1.it

Standardize records of place of death

We suggest that a record of the place of death should be incorporated into death-registration data as a useful additional health metric (*Nature* **494**, 281; 2013).

End-of-life care is a major public-health issue, given the rising number of deaths from chronic illnesses that have multiple and complex symptoms. Knowing where people die can be an indicator of where they were cared for, which is important for allocating health-care resources and for assessing related public-health policies.

Despite surveys that show a prevailing preference for home death among patients, care-givers and the public, most deaths in Europe still occur in hospital (B. Gomes *et al. BMC Palliat. Care* **12**, 7; 2013). We are supplying Portugal with such survey information to help improve the recording of place of death in its electronic death-registration system.

The place of death is registered in a few other countries (the United States and Canada, for example), but its categorization is inconsistent — sometimes even within a country. Location categories need to be internationally standardized (J. Cohen *et al. BMC Public Health* **7**, 283; 2007) and this potentially valuable health-care resource put under the political spotlight.

Barbara Gomes* *King's College London, UK.*

barbara.gomes@kcl.ac.uk

**On behalf of 4 co-signatories (see go.nature.com/jm4tki for full list).*

Nuclear-waste site geology is paramount

As a former geological adviser to the UK government on nuclear-waste repositories, I would like to clarify some points in your discussion of the quest for a British nuclear-waste disposal site (*Nature* **494**, 5–6; 2013).

Nirex was a UK government agency (not an "independent group") that was set up in 1982 to find a geologically suitable site. In 1991, it chose Sellafield in Cumbria — one of two nuclear industry sites — from a list of 537 potentially available locations. Neither of these two sites was among the geologically most suitable, according to Nirex's seven-stage selection process. Its 1997 planning application for an underground laboratory at Longlands Farm, near Sellafield, failed because the inquiry inspector concluded that Nirex did not understand the site's complex geology (see

go.nature.com/5p7yae).

The government's 2008 White Paper, Managing Radioactive Waste Safely (MRWS), put the fact that Cumbria volunteered to consider housing the waste ahead of scientific considerations. This contravenes international guidelines and practice in which national geological searches are conducted before seeking permission from local communities.

To some, this seemed like a back-door attempt to return to the Sellafield district, ignoring both the inspector's original report and the geological problems of the area (see go.nature.com/wob9rf).

You blame a "lack of political will" for the failure of the Nuclear Decommissioning Agency to "sell the facility to local residents". On the contrary, the now-defunct West Cumbria MRWS process spent £3.5 million (US\$5.3 million) on publicity over the past two years.

So Cumbria County Council has demonstrated strong political will by listening to both the geological and the democratic arguments against proceeding with a deep repository for nuclear waste in the region.

David Smythe *University of Glasgow, UK.*

dks1e@udcf.gla.ac.uk

CORRECTIONS

The Outlook article 'A many layered thing' (*Nature* **492**, S52–S54; 2012) contained an error in the graphic 'Caught in a loop'. The labels for the Th1 and Th17 cells were shown switched over.

And in the Outlook article 'Mine, all mine!' (*Nature* **495**, S2–S3; 2013), the map 'Where does gold come from?' originally presented world gold production figures in kilograms but with the label of tonnes. The values have been corrected online to show tonnes.

Robert Richardson

(1937–2013)

Discoverer of superfluidity in helium-3.

Robert Richardson, along with physicist David Lee and myself, discovered that helium-3, a rare but stable isotope of helium, becomes a superfluid when cooled to a minuscule fraction of a degree above absolute zero. Until that discovery, in the early 1970s, the superfluidity effect — in which a liquid flows without friction — had been seen only in helium-4, whose atoms are bosons. No one had been able to produce the phenomenon with fermions, whose nuclei have spin properties different from those seen in bosons.

The discovery astounded the physics community, which had all but given up trying to produce the phenomenon in helium-3. In the decades since, the understanding of superfluidity in this isotope has influenced the development of all sorts of superconducting materials, with applications ranging from power grids to nuclear magnetic resonance (NMR) instruments. It has also helped astrophysicists to better understand neutron stars, which have superfluid cores.

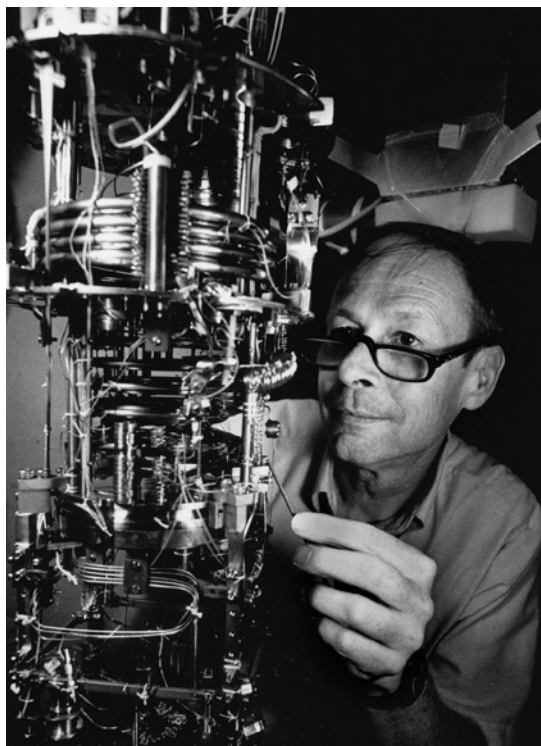
Richardson, who died on 19 February, was born in Washington DC in 1937. He grew up in Arlington, Virginia, where his father worked for a local telephone company. As a boy, Richardson was a keen hiker, camper and birdwatcher. He was a member of the Boy Scouts of America and became an Eagle Scout, the highest rank attainable in the programme, in the shortest time allowed.

Richardson received his undergraduate and master's degrees in physics at Virginia Polytechnic Institute in Blacksburg, before spending six months as an officer in the US Army. Before his military service, he had intended to earn a master's in business administration, but he was unimpressed by the management courses he took in the Army and decided to pursue a PhD in physics instead.

In 1960, Richardson began his PhD work in Horst Meyer's laboratory at Duke University in Durham, North Carolina. It was at Duke that Richardson met Betty McCarthy, who was also a graduate student in physics. The two married in 1962; their daughters, Jennifer and Pamela, were born in 1965 and 1966. In Meyer's lab, Richardson learned how to make the equipment needed to investigate how collections of atoms behave at very low temperatures.

In 1966, soon after Richardson received

his PhD, he and his young family moved to Ithaca, New York, where he became a postdoctoral researcher for David Lee in the low-temperature group at Cornell University. Richardson joined the Cornell faculty in



1968 as an assistant professor in the physics department, and ultimately became a full professor there. I was one of Lee's graduate students when Richardson became a professor. Often working late into the night, Richardson was always around and willing to give advice, irrespective of whose students we were.

By the early 1970s, chemists and materials scientists had for decades been probing the behaviour of materials using NMR — in which nuclei in a magnetic field absorb and re-emit electromagnetic radiation at a specific frequency, depending on the properties of the atoms and the strength of the field. But Richardson played a key part in using the technique for extremely low-temperature studies. The question he, Lee and I set out to answer was whether the direction of the nuclear spins in solid helium-3 could spontaneously order at sufficiently low temperatures.

To produce very low temperatures,

we compressed liquid helium-3 until it began to solidify, causing the mixture of liquid and solid to cool. We initially (mistakenly) believed that we had produced magnetic order in the solid helium-3, at about 0.002 Kelvins. In fact, we had observed a new physical state: superfluidity in liquid helium-3. The three of us shared the 1996 Nobel Prize in Physics for this discovery.

Richardson won or shared many awards and honours apart from the Nobel prize, including the Sir Francis Simon Memorial Prize in 1976 and the Oliver E. Buckley Condensed Matter Prize in 1981. That year, he was also made a fellow of the American Association for the Advancement of Science. He became a member of the US National Academy of Sciences in 1986.

Richardson was an exceptional citizen, at Cornell University and in the broader physics community. From 1990 to 1996, he served as director of the laboratory of atomic-state physics at Cornell. From 1998 to 2007 he was the university's first vice-provost for research. Richardson was also a member of the US National Science Board, which oversees policy issues related to research and education in science and engineering.

Richardson, or Bob to those who knew him, was competitive, passionate and creative, but also humble and considerate. The great sadness of his life was the tragic death, in 1994, of his younger daughter, Pamela, from heart failure. As a way of dealing with their grief, Bob and Betty, together with Alan Giambattista, a physicist at Cornell, threw themselves into writing a physics textbook entitled *College Physics*.

Bob went out of his way to help and guide young scientists. He was an avid participant in 'penny pitching' — a game, played ritually in the low-temperature lab for many years, that involves tossing pennies at a wall to see whose coin lands closest to it — and he often hosted croquet matches at his home. He once challenged a graduate student who had been training for a marathon to a race up several flights of stairs; the student declined, wanting to keep Bob's health intact. ■

Douglas D. Osheroff is professor emeritus at Stanford University, Stanford, California 94305, USA.
e-mail: osheroff@stanford.edu

Persuasive evidence on HIV policy

Antiretroviral therapy has revolutionized the fight against the HIV/AIDS pandemic. Surveillance analyses of a large population in rural South Africa make a compelling case that sustained support for this therapy is essential.

GRACE JOHN-STEWART

In the tragic story of the global HIV/AIDS epidemic, one glimmering light is antiretroviral therapy. This treatment involves a combination of at least three drugs that suppress HIV, and it prevents progression of the infection to AIDS and death. As emergency investment to fight HIV/AIDS evolves into a long-term response, it remains crucial to measure the direct and indirect benefits of HIV treatment to sustain investment. Two studies published in *Science* (Tanser *et al.*¹ and Bor *et al.*²) show a large increase in the overall population life expectancy and a decrease in new cases of HIV infection after expanded use of antiretroviral therapy in South Africa (Fig. 1).

That antiretroviral therapy greatly increases the survival of individuals with HIV has been known since the 1990s³. Global scale-up of this therapy, however, started in 2004, following large investments by the US President's Emergency Plan for AIDS Relief (PEPFAR) and the Global AIDS Program (GAP)⁴. Since then, antiretrovirals have been given to more than 8 million people worldwide (Fig. 2), leading to millions of life-years saved — an unprecedented public-health success story⁵.

Studies of HIV-discordant couples (in which one partner has HIV and the other doesn't) have shown^{6,7} that if the HIV-infected partner receives antiretroviral therapy, transmission rarely occurs. The most compelling confirmation of this observation was a multi-country phase III clinical trial⁸ (HIV Prevention Trials Network 052), in which antiretrovirals decreased the risk of sexual transmission of HIV between couples by 96%. These studies, however, involved motivated patients with high levels of adherence to treatment.

The *Science* papers add a unique perspective on the benefits of antiretroviral therapy on life expectancy and transmission at a general population level. The studies investigate a meticulously characterized population of more than 101,000 individuals from households in a 434-square-kilometre surveillance area in rural KwaZulu-Natal, South Africa. Between 2000 and 2011, the households were surveyed twice a year for birth and death data by the Africa Centre for Health and Population Studies in KwaZulu-Natal. A subset of



Figure 1 | A woman with HIV infection in KwaZulu-Natal, South Africa.

these individuals was serially tested for HIV to estimate the rates of new HIV infections. Starting in 2004, Tanser and colleagues' mapping of individual household location, local HIV prevalence and the proportion of HIV-infected individuals in specified geographical areas who were receiving antiretrovirals (antiretroviral-therapy coverage) allowed them to discern whether living in an area with better coverage was associated with lower risk of acquiring HIV infection.

Both studies demonstrate the profound impact of antiretroviral-therapy coverage at a population level. Bor and colleagues show that expansion of therapy programmes increased the overall adult life expectancy by 11.3 years in KwaZulu-Natal. Tanser *et al.* estimate that in communities in which 30–40% of those infected with HIV were taking antiretrovirals, uninfected individuals were 38% less likely to acquire HIV than those in communities in which the therapy was less widespread.

Population studies such as these, which involve individuals who may have much lower drug adherence than those participating

in carefully controlled clinical trials, are a powerful complement to clinical trials. They give us a sense of impact in the 'real world'. The reality is that the new studies focus on a fairly specific real world — one in which HIV prevalence was staggeringly high, allowing detection of the impact of therapy on overall adult life expectancy. Nonetheless, the studies are persuasive in the larger landscape of global HIV. Several other countries in southern and eastern Africa that also have a high prevalence of HIV could expect to see similar results. In 2012, 25 countries saw a decline of more than 50% in new HIV infections⁵. This decrease was due to multiple factors, but undoubtedly antiretroviral therapy played a pivotal part.

These data, therefore, should persuade policy-makers to sustain or increase investment in this form of therapy. Global aid funds are constricting or staying stable, whereas health needs are persisting or expanding. It is clear from the two studies that for regions of high HIV prevalence, including South Africa, greater investment in antiretroviral

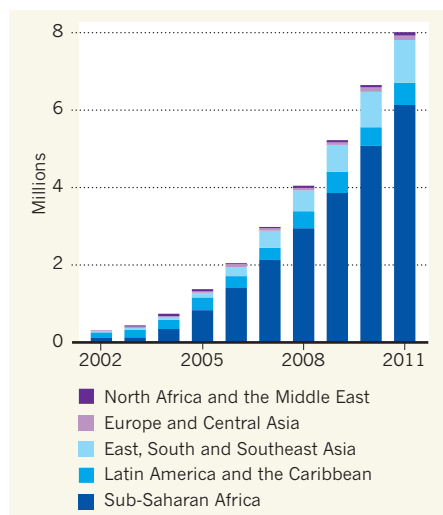


Figure 2 | Use of antiretroviral therapy. The number of people receiving antiretroviral therapy by the end of each year between 2002 and 2011 in low- and middle-income countries by region. (Adapted from ref. 5.)

therapy is essential. To attain one of the goals of the Joint United Nations Program on HIV/AIDS (UNAIDS) — 15 million individuals on antiretroviral therapy by 2015 — continued support from GAP and PEPFAR will be necessary⁵.

Scale-up of antiretroviral therapy has led to gains in programme efficiencies. Not only have drug costs decreased, but also systems to provide antiretroviral therapy have diversified and become less expensive through recruitment of less-highly trained health-care providers and introduction of less-intensive laboratory monitoring⁵. Moreover, funding sources for provision of antiretroviral therapy are branching out, with increased domestic investment by the affected countries to complement external support⁵. For instance, South Africa has substantially increased its domestic investment in antiretroviral therapy, and the present studies provide clear justification for sustaining and expanding this investment.

At a national or global policy level, antiretroviral therapy has been the subject of scrutiny in many ways: for its dual treatment and prevention benefits; for its impact on other health needs such as childhood illnesses, reproductive health or other infectious diseases, including tuberculosis; for the potential strain the therapy places on health systems, for example whether there is an uneven focus on HIV/AIDS at the expense of other illnesses in settings with shortages of health workers and resources; and for its contribution to well-being or economic productivity. Antiretroviral-therapy programmes have led to broad improvements in health systems, innovations in clinical education and overhauls of laboratory and clinical infrastructure — much more than the global community initially envisioned. Financial investments

were linked to measurable and diverse beneficial outcomes.

For the future, it remains important to wrestle with what outcomes are most relevant to health-resource investment decisions. Although increased overall life expectancy and a reduction in new HIV infections at a population level are compelling for policy-makers in regions with high HIV prevalence, other metrics, including comparisons with other health investments and impact on morbidity, quality of life or productivity, may be more useful in lower-prevalence settings. ■

Grace John-Stewart is in the Department of Global Health, University of Washington,

Seattle, Washington 98104, USA.

e-mail: gjohn@uw.edu

1. Tanser, F., Barnighausen, T., Grapsa, E., Zaidi, J. & Newell, M. L. *Science* **339**, 966–971 (2013).
2. Bor, J., Herbst, A. J., Newell, M. L. & Barnighausen, T. *Science* **339**, 961–965 (2013).
3. Palella, F. J. Jr et al. *N. Engl. J. Med.* **338**, 853–860 (1998).
4. PEPFAR Blueprint: Creating an AIDS-Free Generation (US State Dept, 2012); available at go.nature.com/h27z18
5. *Global Report: UNAIDS Report on the Global AIDS Epidemic 2012* (UNAIDS, 2012); available at go.nature.com/c4lupm
6. Bunnell, R. et al. *AIDS* **20**, 85–92 (2006).
7. Donnell, D. et al. *Lancet* **375**, 2092–2098 (2010).
8. Cohen, M. S. et al. *N. Engl. J. Med.* **365**, 493–505 (2011).

MATERIALS SCIENCE

Nanoparticle structures served up on a tray

A neat approach that involves laying an array of nanoparticles on a graphene sheet supported on an iridium substrate has allowed accurate measurement of the nanoparticles' atomic structure.

SIMON BILLINGE

Whether we are aware of the fact or not, modern life would be profoundly different without precious-metal nanoparticle catalysts¹. Every drop of petrol used has passed over a platinum catalyst during production, and the exhaust gases of a car's engine similarly pass over these catalysts in the catalytic converter on the way out of the car. The catalysts are also essential to future sustainable energy technologies, such as cars and buses powered by hydrogen fuel cells. These processes rely on chemical reactions that take place on the catalyst surface. But despite our dependence on such reactions, we know little about them because we lack the tools that can 'see' the processes involved and measure the interactions between reacting molecules and the catalyst. Writing in *Physical Review Letters*, Franz et al.² describe an ingenious exploitation of a gift from nature that will allow us to make a huge dent in this problem.

A catalyst plays host to a chemical reaction, allowing it to proceed without itself being consumed in the process. Much like a skilled host at a party, it brings together like-minded guests who then depart, leaving the host to find and facilitate other favourable matchings. Understanding the secrets of the successful matchmaker will help us to find better, cheaper and more naturally abundant materials that do the same job. But how do we unlock those secrets?

Catalysis by precious metals takes place on the surface of minute nanoparticles, which maximize the surface area available for the chemical reactions and, equally importantly, minimize the amount of expensive precious metal needed.

To understand these catalytic chemical processes, we must follow the interactions between the reagents and the metal atoms on the nanoparticle surface. The surface atoms become rearranged to minimize the surface energy, a feature called surface reconstruction, and the reactivity and interactions between the atoms and the reagents vary widely depending on this rearrangement. But not only do we not know what the surface structure looks like, we do not even have robust methods for working out the 'bulk' nanoparticle structure, a dilemma sometimes referred to as the nanostructure problem³.

When atoms become arranged into a periodic structure, as if they were on a chessboard, they form crystals and we have powerful tools for locating them. More recently, with the advent of intense synchrotron X-ray sources based on particle accelerators, these methods have been extended to the study of surface atoms. But, crucially, this can be done only when the arrangement of the surface structure is also periodic. When the arrangement exists only on the nanoscale it is invisible, as if we are blindfolded. There are two approaches to addressing this problem: either develop tools for solving structure at the nanoscale⁴,

or arrange the nanoparticles into ordered periodic arrays and use the existing tools of crystallography⁵.

Franz and colleagues' ingenious approach is to serve up the nanoparticles on a tray made of graphene (a carbon sheet just one atom thick) supported on an atomically flat iridium surface. The gift from nature is that the delicate mismatch between the iridium substrate and the graphene tray produces an orderly array of ripples in the graphene called a moiré pattern. This array is the perfect size in which to nestle identical 82-atom nanoparticles of iridium⁶ arranged periodically and with the same orientation (Fig. 1) — exactly the requirements for solving a crystallographic structure. For the first time, our blindfold is removed and accurate quantitative details about the nanoparticles are revealed, including the nature of their surfaces and the modification to their structure that results from their interactions with the substrate.

These results would have far-reaching consequences for the understanding of catalysis if the system could be studied when it is undergoing catalysis and if the changes in structure and surface chemistry were followed under different conditions, such as varying temperature. Powerful theoretical approaches for following these reactions are also available, but these have never been validated against experiment because of the lack of results from atomically resolved experiments such as this. That is the good news. The bad news is that nature gifted us with the structure of iridium nanoparticles and not with that of the industrially used platinum, which lies tantalizingly one place above iridium in the periodic table.

The authors hope that the magic can be made to work with platinum as well, but this remains to be seen. Delicate tuning of the metal atoms and substrate is required for them to form periodically arranged nanoparticles: too good a match, and a completely uniform monolayer will form; too poor a fit, and they will not remain in registry with the moiré pattern and will not be periodic and orientationally aligned. There is still a lot that can be learned from this model system of iridium particles. However, we can expect a much greater impact if the trick can be repeated, not just with platinum nanoparticles, but also with a range of materials, so that catalytic performance can be correlated with changes in structural and electronic properties.

Even so, commercial catalysts are not made in this way, and a final understanding of what is happening in a commercial catalytic stack will require study of the nanoscale structure *in situ* in a temporally and spatially resolved way in a complex, heterogeneous system. This is the frontier of complexity that must be pushed back if the high-performance materials needed in many next-generation technologies for energy, the environment and health are to be developed. It will take a combination

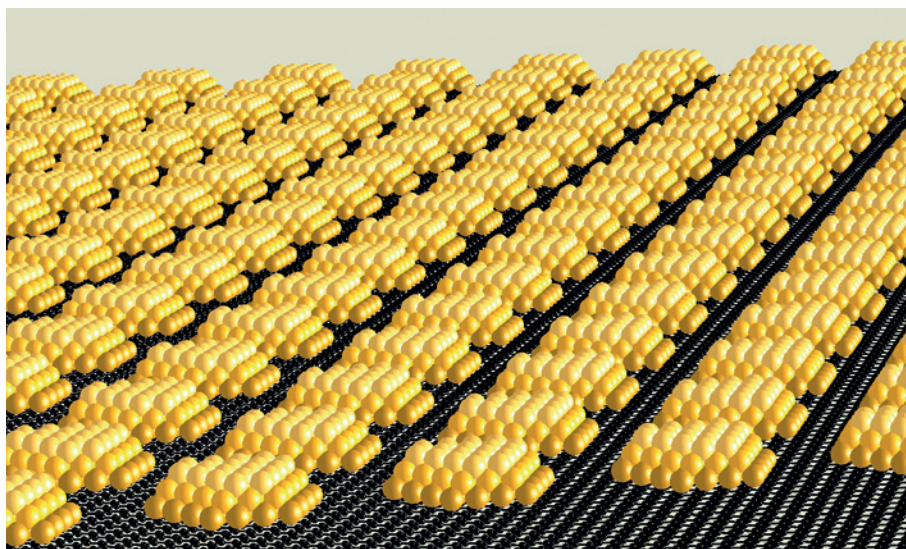


Figure 1 | Nestling nanoparticles. Franz *et al.*² have measured the atomic structure of 82-atom nanoparticles of iridium (yellow) by arranging them periodically in a graphene–iridium surface (black) and applying surface X-ray crystallography.

of the most powerful emerging experimental tools, validated computational methods, ingenuity and further gifts from nature. But we are close to being able to do this, and steps such as those reported by Franz *et al.* take us even closer. ■

Simon Billinge is in the Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA, and in the Department of Condensed Matter Physics and Materials Science, Brookhaven National

Laboratory, Upton, New York.
e-mail: sb2896@columbia.edu

1. Gates, B. C. *Chem. Rev.* **95**, 511–522 (1995).
2. Franz, D. *et al. Phys. Rev. Lett.* **110**, 065503 (2013).
3. Billinge, S. J. L. & Levin, I. *Science* **316**, 561–565 (2007).
4. Juhás, P., Cherba, D. M., Duxbury, P. M., Punch, W. F. & Billinge, S. J. L. *Nature* **440**, 655–658 (2006).
5. Jadzinsky, P. D., Calero, G., Ackerson, C. J., Bushnell, D. A. & Kornberg, R. D. *Science* **318**, 430–433 (2007).
6. N'Diaye, A. T., Bleikamp, S., Feibelman, P. J. & Michely, T. *Phys. Rev. Lett.* **97**, 215501 (2006).

EARTH SCIENCE

How the ocean exhales

A record of biogenic opal production in the subtropical Atlantic Ocean fuels the theory that ocean circulation, rather than winds, drove the release of carbon dioxide from deep marine waters at the end of the last ice age. [SEE LETTER P.495](#)

ELISABETH SIKES

The concentration of atmospheric carbon dioxide was lower during the last ice age¹ than in the pre-industrial modern age, and it is widely accepted that changes to deep-ocean circulation during this glacial period increased the amount of CO₂ sequestered in the deep ocean. When deep-ocean circulation shifted to modern patterns, the sequestered CO₂ was vented back to the atmosphere, most probably through the Southern Ocean. Two plausible controlling mechanisms for this degassing have been proposed, one in which latitudinal shifts in winds trigger CO₂ release and synchronize post-glacial warming

in the Northern and Southern hemispheres², and another in which ocean circulation triggers venting and so provides the inter-hemispheric link³. On page 495 of this issue, Meckler *et al.*⁴ report evidence of enhanced growth of silica-producing (siliceous) plankton known as diatoms (Fig. 1) at low latitudes at the time when CO₂ was rapidly venting. This suggests that the degassing was triggered by transient changes in deep-water density driven by fluctuations in northern-water masses, rather than by a shift in Southern Ocean winds.

Although the ocean surface is largely in equilibrium with the atmosphere, the deep ocean holds 50 times more CO₂ than the

atmosphere. This is because of a process called the biological pump: first, phytoplankton fix CO_2 in the ocean surface by photosynthesis to form organic carbon; then, when they die, the phytoplankton sink into the deep ocean, where the organic carbon is consumed by other organisms and respired back to CO_2 , which forms dissolved carbonate.

The biological pump also depletes the surface waters of nutrients such as nitrogen (needed by all phytoplankton) and silica (needed by diatoms), transporting them to the deep ocean along with the carbon. These nutrients and dissolved carbonate become trapped in the deep ocean by the shallow surface layer, which acts somewhat like the lid on a bottle of a fizzy drink. Ocean circulation returns water containing high levels of carbonate and silica to the surface in the Southern Ocean and at upwelling regions. When this happens, biological productivity at the surface is enhanced and — if degassing outpaces the biological pump — CO_2 is released to the atmosphere. A record of the elevated productivity is preserved when the siliceous remains of the organisms settle to the sea floor and accumulate as opal sediments.

During glaciations, factors such as reduced formation of deep water in the North Atlantic and increased ice cover capping the Southern Ocean effectively trapped more CO_2 in the ocean. Early in the most recent deglaciation period, iceberg discharge into the North Atlantic shut down the formation of deep waters in this region⁵. This fundamental change in the organization of deep-water masses brought warmth to the Southern Ocean and could have enhanced upwelling, CO_2 release and southern deep-water formation because of

density changes alone³. The warming has also been suggested to have triggered a shift of the westerlies towards Antarctica, forming part of a feedback mechanism in which the shifted winds caused upwelling and more CO_2 to be vented up to the atmosphere, leading to yet more warming, wind changes and CO_2 release². Conversely, evidence from opal deposition in the Southern Ocean^{6,7} suggests that the core of the southern westerly winds shifted to the south early in the deglaciation period and that this may have been the main cause of invigorated CO_2 release.

Meckler *et al.* follow the trail of opal deposition to the subtropical North Atlantic, and build on the concept⁸ that increased opal sedimentation at low latitudes reflects stimulated diatom productivity caused by an increase in northward transport of dissolved silicate from the Southern Ocean. They observed evidence of simultaneous rapid opal sedimentation in both the subtropical North Atlantic and the Southern Ocean during the deglacial shut-down of northern deep-water formation, and depart from the prevailing interpretation by inferring that this corresponded to a slowdown in transport of subsurface water from the south. The authors correctly point out that the combination of stronger winds and enhanced diatom productivity around Antarctica would have increased the formation of silicate-poor water in the south; because this water feeds shallow subsurface waters (the thermocline) at low latitudes, it would have depressed productivity in the tropical North Atlantic.

The authors instead suggest that reduced formation of waters at intermediate depths in both hemispheres caused a shallowing of the boundary of the silicate-rich southern deep

waters (see Fig. 4b of the paper⁴). This would have allowed the transfer of silicate and dissolved carbon from the deep waters into the photic zone of the ocean — the upper ocean region that receives enough light for photosynthesis to occur. Similar to the situation in the Southern Ocean today, a competition would have ensued in which enhanced productivity in the photic zone pulled CO_2 back down into deep waters, even as more dissolved carbon and nutrients emerged from the deep. Meckler and colleagues propose that the flow of CO_2 and nutrients outpaced the biological pump, and conclude that the observed opal sedimentation is diagnostic of deep-ocean reorganization and carbon out-gassing during deglaciation, rather than the signature of carbon draw-down.

The prevailing consensus has shifted towards winds in the Southern Hemisphere as being the driver of CO_2 release during the last deglaciation^{6,7}. Meckler *et al.* put the ocean back in control and keep the trigger in the Northern Hemisphere. The authors' case hinges, in part, on a re-evaluation of the implications for ocean circulation of a range of geochemical evidence dating from that crucial time period. This evidence was previously proposed to indicate an increased influx of silicate-rich shallow subsurface waters to the low-latitude Atlantic. Meckler and colleagues attribute the evidence to shallowing of a silicate-rich deep-water mass (as detailed in the Supplementary Information for their paper⁴). Further understanding will come with a refinement of the timing of key deglacial events^{9,10} and with a fuller characterization of the shallow water masses that form at high latitudes and which feed the thermocline at low latitudes¹¹. The authors' study provides a valuable piece of the puzzle for deciphering the ocean's role in driving CO_2 changes during glacial and deglacial periods. ■

Elisabeth Sikes is at the Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901-8521, USA.

e-mail: sikes@marine.rutgers.edu

- Monnin, E. *et al. Science* **291**, 112–114 (2001).
- Toggweiler, J. R., Russell, J. L. & Carson, S. R. *Paleoceanography* **21**, <http://dx.doi.org/10.1029/2005PA001154> (2006).
- Broecker, W. S. *Paleoceanography* **13**, 119–121 (1998).
- Meckler, A. N. *et al. Nature* **495**, 495–498 (2013).
- McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. *Nature* **428**, 834–837 (2004).
- Anderson, R. F. *et al. Science* **323**, 1441–1448 (2009).
- Toggweiler, J. R. & Lea, D. W. *Paleoceanography* **25**, <http://dx.doi.org/10.1029/2009PA001758> (2010).
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. *Nature* **427**, 56–60 (2004).
- Shakun, J. D. *et al. Nature* **484**, 49–54 (2012).
- Burke, A. & Robinson, L. F. *Science* **335**, 557–561 (2012).
- Rose, K. A. *et al. Nature* **466**, 1093–1097 (2010).



Figure 1 | A source of silica. Diatoms are a major group of phytoplankton that encase themselves in a silica shell. When these marine organisms die, their shells sink to the ocean floor, forming opal sediments. Meckler *et al.*⁴ have used opal sediments as a geochemical record of ocean circulation at the end of the last ice age.

X-RAY CRYSTALLOGRAPHY

One size fits most

Crystalline 'sponges' offer a way to impose order on small molecules so that their structures can be solved by X-ray crystallography. This enables nanogram quantities of material to be analysed using the technique. [SEE ARTICLE P.461](#)

PIERRE STALLFORTH & JON CLARDY

Much of science involves describing things that we cannot see, and no method has revealed more about the hidden world of molecular structures than X-ray crystallography¹. But the sample preparation required for this technique can be difficult, especially when only small quantities of a compound are available. On page 461 of this issue, Inokuma *et al.*² report an approach that can overcome this problem, greatly increasing the technique's applicability.

X-ray crystallography has been crucial for solving the structures of an array of small molecules, including antibiotics, industrial catalysts, artificial sweeteners and the neurotransmitters that act as signals in the brain. The importance of the technique is reflected by the effect that a paper on the subject can have on journal impact factors. For example, when an article³ describing the most widely used set of programs for analysing small-molecule X-ray data was published in 2008, the impact factor of the publishing journal jumped from 2.05 in the year of publication, to 49.93 in 2009 and 54.33 in 2010, sinking back to 2.08 in 2011.

Given the prominence of the method, it is surprising to discover that X-ray crystallography has a major liability: it requires crystals.

The regular, repeating arrangement of molecules in crystals scatters X-rays into a series of amplified and spatially constrained beams whose intensity can be easily measured. The intensities are computationally processed to provide an image of the molecule in the crystal. Single molecules, however, scatter X-rays to produce a weak, continuous beam, which is of little use for analytical purposes.

Of course, other techniques such as mass spectrometry and nuclear magnetic resonance can be used to obtain molecular structures, but they do so through an inferential process of ruling out alternatives^{4,5} that differs fundamentally from the images generated by X-ray crystallographic analysis. Nevertheless, these alternative techniques can work with sample sizes much smaller than those required for X-ray crystallography, and, most importantly, they do not require crystals⁶.

X-ray analysis has benefited from the highly intense X-ray beams produced by synchrotrons, from the development of efficient ways for measuring scattered beams and from powerful algorithms used for the computational processing that provides molecular images. These advances directly or indirectly allow smaller crystals to be used, but none of them addresses the need for crystals. Growing crystals remains more of an artisanal skill than an intellectual

one, and is based on few general principles and lots of trial and error. These difficulties are often compounded by the limited quantities available for many samples — in general, it is difficult to crystallize samples of less than one milligram.

Inokuma *et al.* report a way to circumvent crystallization of the sample. Their approach is breathtakingly simple: they grow host crystals that have large, well-defined, interconnected cavities, and allow these crystals to soak up small molecules from a solution of the sample. If the small molecules are the right size and shape to be trapped by the host's cavities, they become arranged in an orderly fashion that allows them to be studied using X-ray crystallography.

By way of analogy, imagine a departure gate at a busy airport, with seats arranged in a two-dimensional grid. Passengers for the flight that departs from this gate enter the airport through different entrances, mingle with passengers for many other flights and explore different paths and attractions on their way to the gate. But shortly before departure, only the passengers on the departing flight occupy the seats, or most of the seats, in the lounge. In this analogy, the empty, regularly arranged seats are the cavities in the host crystals, and the selection principle — the appropriate boarding pass — is molecular shape.

For host crystals, Inokuma and colleagues used a type of material that has many names: metal-organic frameworks, porous materials, or, perhaps most appropriately for this application, crystalline sponges⁷. The structures of these sponges are nothing like the illustrations of crystals found in introductory texts, which depict spheres packed tightly together — similar to the elements of tessellation paintings by M. C. Escher (Fig. 1a). Instead, they resemble the steel scaffolding of skyscrapers under construction, with the large internal cavities of

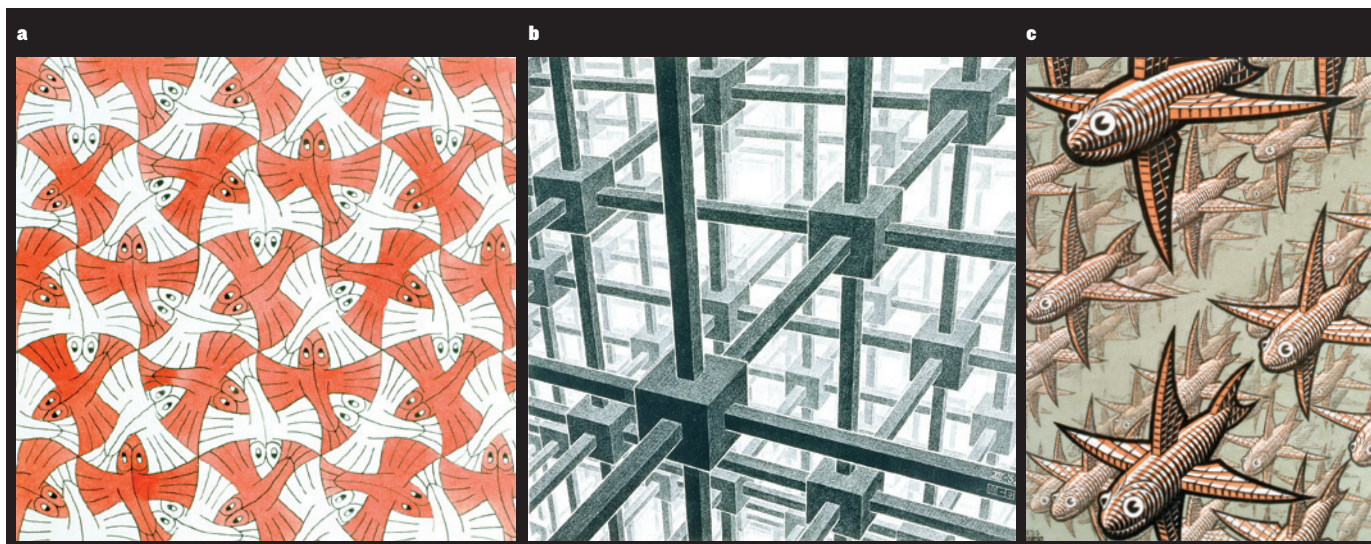


Figure 1 | Order imposed by a framework. These paintings by M. C. Escher exemplify features of different crystal structures. **a**, In most crystals, the molecules are tightly packed together, as are the fish in *Symmetry Drawing*. **b**, The components of crystalline sponges, however, adopt a regular porous

network, such as the one shown in *Cubic Space Division*. **c**, Inokuma *et al.*² induced a lattice-like arrangement of molecules by trapping them in the pores of crystalline sponges, much like the arrangement of fish in *Depth*. This enabled the authors to solve X-ray crystal structures of nanogram quantities of compounds.

future rooms outlined by steel girders (Fig. 1b). In crystalline sponges, the girder intersections are metal atoms, and the girders are long, linear, rigid organic molecules. The properties of the internal cavities depend on the properties, such as the length, of the connecting groups.

It was known that crystalline sponges could be designed to absorb molecules selectively, but the authors report variants that have just the right degree of promiscuity (the opposite of selectivity). The materials are promiscuous enough to accommodate many different shapes of molecules, but selective enough to hold trapped molecules in only one orientation. The sponges also confine the molecules loosely enough to allow reversible binding, so that the molecule can try out several different binding orientations until it finds the one with the lowest energy. The researchers found that their crystalline sponges pulled most of the guest molecules they tested out of solution and held them in an ordered array (Fig. 1c) so that clear structures of the molecules could be obtained using X-ray crystallography.

Inokuma and colleagues used their method to solve the structures of an impressively diverse set of examples, including a naturally occurring molecule, miyakosyne A, that had defied full analysis by other methods. One of the researchers was also able to determine the structures of six molecules in a blind test. Impressively, the technique reduces the amount of sample needed for analysis by up to about 1,000-fold, compared with quantities that are typically required for conventional X-ray crystallography.

It remains to be seen what percentage of molecules will be amenable to the authors' technique. Moreover, molecules must fit into a single cavity in the sponge: miyakosyne A seems to be near the upper size limit for which the reported sponges will be useful. And for the method to be widely adopted, the crystalline sponges developed by Inokuma and co-workers must become easily available, for example from a commercial supplier. But if these initial results are generally applicable, one can expect to see an increase both in reported

molecular structures and in the preparation of alternative crystalline sponges. One can even imagine that, in the near future, researchers will not bother trying to crystallize new molecules. Instead, a small set of crystalline sponges will be used to absorb the molecules, and X-ray analyses performed to see which sponge allows the clearest image of the molecule. ■

Pierre Stallforth and Jon Clardy are in the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. e-mail: jon_clardy@hms.harvard.edu

1. Blow, D. *Outline of Crystallography for Biologists* (Oxford Univ. Press, 2002).
2. Inokuma, Y. *et al.* *Nature* **495**, 461–466 (2013).
3. Sheldrick, G. M. *Acta Crystallogr. A* **64**, 112–122 (2008).
4. Nuzillard, J.-M. & Massiot, G. *Tetrahedron* **47**, 3655–3664 (1991).
5. Lederberg, J. *et al.* *J. Am. Chem. Soc.* **91**, 2973–2976 (1969).
6. Molinski, T. F. *Nat. Prod. Rep.* **27**, 321–329 (2010).
7. Kitagawa, S., Kitaura, R. & Noro, S.-I. *Angew. Chem. Int. Edn* **43**, 2334–2375 (2004).

CELL BIOLOGY

Receptor signals come in waves

The activity of G-protein-coupled receptors is not limited to the cell surface. Evidence from microscopy points to three temporally, spatially and perhaps functionally distinct waves of signalling by these receptors. SEE LETTER P.534

MARTIN J. LOHSE & DAVIDE CALEBIRO

Nanobodies, as their name suggests, are the smallest fully functional antibody fragments known¹. They represent the end domain of single-protein-chain antibodies that occur naturally in camels and a few other animals². Because of their small size and their flexibility, which allows access to parts of proteins that normal antibodies cannot reach, they have become of great interest to research into biological receptors. In fact, they have been instrumental in stabilizing and crystallizing the active forms of G-protein-coupled receptors (GPCRs) and their associated G proteins³, which are of relevance to the treatment of many disorders, including heart disease, asthma, allergies and pain. In this issue (page 534), Irannejad *et al.*⁴ use such active-state-specific nanobodies to determine where and when GPCRs and their G proteins signal in cells*.

The conventional view of GPCR signalling holds that the receptor couples with its

G protein only at the cell membrane. Several events are then thought to terminate this signalling, including binding of the inhibitor proteins β -arrestins⁵ to the receptor and internalization of the receptor through the process of endocytosis in clathrin-coated pits, which pinch off the membrane to form intracellular vesicles called endosomes. It was subsequently discovered⁶ that β -arrestin binding to a GPCR may itself trigger a distinct wave of 'non-conventional' signals downstream, such as activation of ERK enzymes, which regulate cell growth and survival.

Once inside endosomes, cell-surface receptors are either degraded or recycled back to the cell membrane; they were, however, thought to be functionally inactive. Nonetheless, it emerged that receptor signalling may also occur from endosomes⁷. Indeed, such internal signalling to G proteins has been observed for the GPCRs of parathyroid hormone⁸ and thyroid-stimulating hormone (TSH)⁹, and of the signalling molecules sphingosine-1-phosphate¹⁰ and dopamine¹¹. This activity of internalized receptors is thought to

constitute a third wave of signalling, following the G-protein-mediated and β -arrestin-mediated waves.

Irannejad *et al.* now provide photographic and video evidence for all three signalling waves. They investigated the well-studied GPCR known as the β_2 -adrenergic receptor (β_2 AR) and its associated G_s protein. β_2 AR- G_s activity has been shown to stimulate production of the intracellular second-messenger molecule cyclic AMP in a reaction catalysed by the enzyme adenylyl cyclase. This in turn induces physiological responses such as dilation of blood vessels and bronchi, and the tremor that comes with excitement or fear. Irannejad and co-workers used two nanobodies: Nb80, which binds to active β_2 AR, and Nb37, which recognizes guanine-nucleotide-free G_s — a transition state in G_s activation.

The authors detected two temporally and spatially separated waves of β_2 AR signalling through G_s . Nb80 was recruited to the active receptor not only at the cell membrane but also — albeit later on — on the membrane of intracellular vesicles. Similarly, active G_s was detected with Nb37, both at the cell membrane and, later on, on intracellular vesicles. Both waves of signalling led to the accumulation of cAMP.

Between these two waves, β_2 AR bound to β -arrestin and then clustered in clathrin-coated pits, where it is thought to trigger a wave of non-conventional signals, including ERK activation (Fig. 1). Intriguingly, neither Nb80 nor Nb37 bound to the pits, suggesting that both Nb80 and G_s dissociated from β_2 AR at this stage. It could be that the β -arrestins actively displace the nanobodies, or that only the receptors that are not labelled

*This article and the paper under discussion⁴ were published online on 20 March 2013.

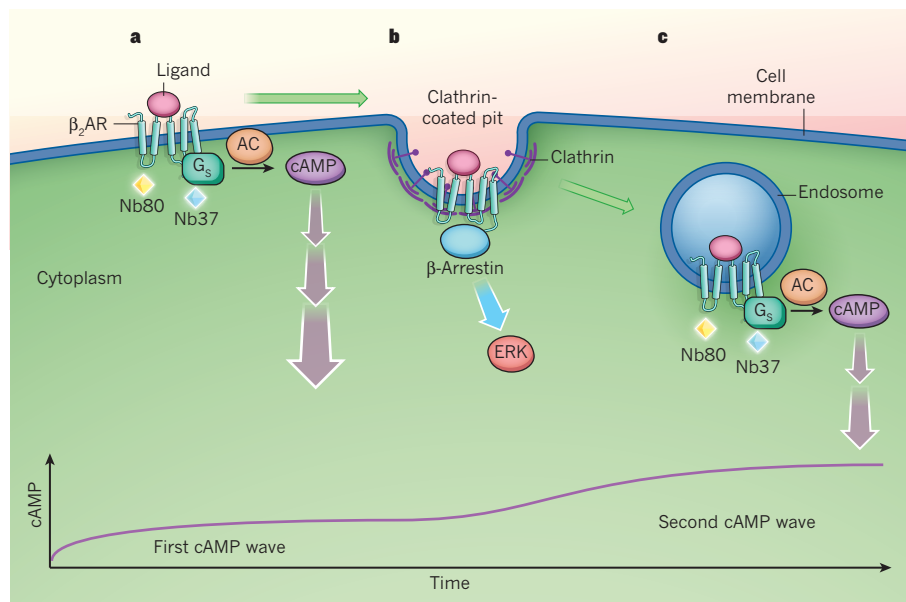


Figure 1 | One receptor, three signals. Irannejad *et al.*⁴ show three waves of signalling by the G-protein-coupled β_2 -adrenergic receptor (β_2 AR). **a**, The first wave occurs on the cell membrane and involves ligand–receptor interaction with the G_s protein, resulting in increased production of cAMP molecules by the enzyme adenylyl cyclase (first cAMP wave; see graph). **b**, The second wave occurs from clathrin-coated pits (and possibly their downstream vesicles) and involves receptor interaction with β -arrestin proteins, activation of ERK enzymes and other non-conventional signals. **c**, Wave three occurs after receptor internalization to intracellular vesicles — most probably endosomes — and is the second wave to involve G_s , resulting in further increases in intracellular cAMP (second cAMP wave). The authors detected active forms of β_2 AR and G_s with the Nb80 and Nb37 nanobodies, respectively.

by Nb80 reach the clathrin-coated pits.

Cells abound with various types of intracellular vesicle, and it should be determined in which of these the second G-protein-mediated signalling wave occurs. Work on parathyroid-hormone receptors^{8,12}, and now on β_2 AR, points to early endosomes, whereas studies of TSH receptors^{9,13} suggest that the vesicles in which they were signalling belonged to the Golgi complex or *trans*-Golgi network, which mediate intracellular trafficking. It could be that receptor-containing endosomes eventually fuse with other vesicular structures where the receptors might meet ‘new’ G proteins.

It would be interesting to know whether and how ligands that activate GPCRs are internalized with the receptors to produce the second wave of cAMP production. When such ligands bind to receptors with high affinity (as in the case of TSH and parathyroid hormone), they could simply remain bound to the receptors during internalization. For low-affinity, rapidly dissociating ligand–receptor pairs, such as β_2 AR and its ligand, this is unlikely. Nevertheless, a certain number of dissociated ligands might be internalized in the same vesicle as their receptor, and in the tight space within the vesicle the two molecules might reunite.

Strong evidence indicates that GPCRs can exist in several active conformations^{14,15}. Are the three waves of β_2 AR signalling associated with different conformations of this receptor? Notably, Nb80 seems to dissociate from β_2 AR before the receptor enters clathrin-coated pits, which could be because β_2 AR adopts a

different active conformation when it binds to β -arrestin.

It will be particularly exciting to discover whether the two distinct waves of G-protein-mediated signalling have different physiological consequences, and how to specifically interfere with each for therapeutic purposes. It seems that cAMP generated at the cell membrane and at internal sites in the cell may

produce different responses. In the case of TSH-receptor signalling, for example, the stimulation of thyroid-hormone secretion has been hypothesized to involve cAMP generated internally⁹; this second wave of cAMP signalling, and the secretion response, can be inhibited by blocking receptor internalization^{4,9}.

Irannejad and colleagues’ results will not just entice researchers to dig deeper. They also show once more that nature has invented many tricks and dodges that allow a few chemical messages to transmit a wealth of biological information into a cell. ■

Martin J. Lohse and Davide Calebiro are at the Institute of Pharmacology and Toxicology, and the Rudolf Virchow Center, University of Würzburg, 97078 Würzburg, Germany. e-mail: lohse@toxi.uni-wuerzburg.de

1. Harmsen, M. M. & De Haard, H. J. *Appl. Microbiol. Biotechnol.* **77**, 13–22 (2007).
2. Hamers-Casterman, C. *et al. Nature* **363**, 446–448 (1993).
3. Steyaert, J. & Kobilka, B. K. *Curr. Opin. Struct. Biol.* **21**, 567–572 (2011).
4. Irannejad, R. *et al. Nature* **495**, 534–538 (2013).
5. Lohse, M. J., Benovic, J. L., Codina, J., Caron, M. G. & Lefkowitz, R. J. *Science* **248**, 1547–1550 (1990).
6. Shenoy, S. K. & Lefkowitz, R. J. *Trends Pharmacol. Sci.* **32**, 521–533 (2011).
7. Calebiro, D., Nikolaev, V. O., Persani, L. & Lohse, M. J. *Trends Pharmacol. Sci.* **31**, 221–228 (2010).
8. Ferrandon, S. *et al. Nature Chem. Biol.* **5**, 734–742 (2009).
9. Calebiro, D. *et al. PLoS Biol.* **7**, e1000172 (2009).
10. Mullershausen, F. *et al. Nature Chem. Biol.* **5**, 428–434 (2009).
11. Kotowski, S. J., Hopf, F. W., Seif, T., Bonci, A. & von Zastrow, M. *Neuron* **71**, 278–290 (2011).
12. Feinstein, T. N. *et al. Nature Chem. Biol.* **7**, 278–284 (2011).
13. Werthmann, R. *et al. FASEB J.* **26**, 2043–2048 (2012).
14. Zürn, A. *et al. Mol. Pharmacol.* **75**, 534–541 (2009).
15. Nygaard, R. *et al. Cell* **152**, 532–542 (2013).

PALAEONTOLOGY

Tubular worms from the Burgess Shale

The discovery, in 500-million-year-old rocks, of fossil acorn worms that lived in tubes illuminates the debate about whether the ancestor of vertebrates was a mobile worm-like animal or a sessile colony dweller. [SEE LETTER P.503](#)

HENRY GEE

Understanding the origin of vertebrates is fascinating not just because it appeases our self-interest, but because it is inherently difficult. Vertebrates seem to have a qualitatively more complex construction than other animals. This means not only that the origin of distinctive features of vertebrate anatomy, such as the head or

the neural crest, are difficult to trace among invertebrates, but also that the evolutionary roots of vertebrates as a whole are hard to fathom. On page 503 of this issue, Caron *et al.*¹ address a fundamental issue in this regard: whether the ancestor of the deuterostomes (the larger assemblage to which vertebrates belong) was a free-living, worm-like creature or a sessile, perhaps colonial, animal. Their answer, for now, is that it was

solitary, worm-like and motile*.

Caron and colleagues describe fossilized marine worms found in the Burgess Shale of western Canada, which dates to 505 million years ago, in the Middle Cambrian period. The fossils — which the authors name *Spartobranchus tenuis* — are enteropneusts, also known as acorn worms, and the authors show that these particular worms are unusual in that they lived at least part of their lives in tubes. This dwelling habit is otherwise unknown in enteropneusts, but is common in their minute, typically colonial cousins, the pterobranchs. At first glance, the discovery seems modest enough, but the finding helps to resolve the sometimes vexed question of the relationships between enteropneusts and pterobranchs and, from that, the probable habit of the earliest deuterostomes (Fig. 1).

The anatomy of enteropneusts has long fascinated those questing to understand vertebrate origins. The early career of the biologist William Bateson was spent studying these creatures to that end. However, the lack of a fossil record, and any sound way to reconstruct phylogeny, thwarted his attempts — instead, he went off to become the founder of genetics².

*This article and the paper under discussion¹ were published online on 13 March 2013.

What makes enteropneusts especially interesting is that their midsection is perforated by a series of openings, the pharyngeal gill slits. These seem to be distinguishing features of deuterostomes as a whole: they are found in chordates (including some vertebrates) and arguably in some primitive echinoderms (starfish and their allies). Enteropneusts have hints of other features that suggest a close alliance with chordates, such as a dorsal, hollow nerve cord and an axial stiffening rod called a notochord. These features, if they occur at all, are now thought to have evolved in parallel in these clades. But the occurrence of pharyngeal slits seems to be a solid benchmark of common ancestry.

Molecular work over the past 20 years or so has revolutionized our understanding of deuterostome relationships. It is now clear that enteropneusts and pterobranchs (which are collectively referred to as the hemichordates) are more closely related to echinoderms than to other deuterostomes³. This discovery revitalized and added weight to a long-held, but contentious, suggestion that serial structures in some primitive echinoderms might be pharyngeal gill slits homologous with those found in chordates^{4,5}.

Another focus of interest is the relationship between enteropneusts and pterobranchs. There is no doubt that they are closely related,

but which came first, the generally sessile lifestyle of pterobranchs or the free-living mode seen in enteropneusts? Because hemichordates generally seem rather primitive, the probable tube-dwelling habit of the deuterostome ancestor hangs on this question. Most early researchers backed one idea or the other, although zoologist Walter Garstang, then seen as a pre-eminent authority, backed both. His initial work suggested a motile ancestry⁶, but later in his career he settled on sessility⁷. However, more recent work has suggested something rather startling: that the habit of the enteropneust is the more primitive of the two, and that pterobranchs are most closely related to a particular group of enteropneusts — the harrimaniids⁸. This finding meant that pterobranchs could be considered as somewhat specialized, sessile versions of enteropneusts.

However, the question remained: how did a large, free-living worm evolve into the minuscule tube-dwelling pterobranch that filters detritus from sea water using an elaborate crown of tentacles. The fossil record preserves all kinds of possibly intermediate morphologies between one creature and another. But the record is notoriously delphic, if not laconic, and on this subject it has even less to say than usual. Enteropneusts, being soft-bodied, do not preserve well as fossils: their record in stone is exiguous in the extreme. Before this report, the earliest-known fossil enteropneusts were Triassic, between 250 million years and 200 million years old. That Caron *et al.* extend the fossil record of enteropneusts back to the Cambrian period makes their findings notable, even with no other contribution.

The horny tubes of pterobranchs, by contrast, preserve rather well. Amateur geologists will know them as graptolites. The animals that live inside the tubes, however, are scarcely represented in the fossil record. And this is why the specimens described by Caron *et al.* are so important. They show early enteropneusts that are very similar to modern harrimaniids, but that are clearly associated with pterobranch-like tubes. This suggests that the ancestor of pterobranchs adopted a tube-living habit before becoming colonial and sessile, and before evolving elaborate tentacles (Fig. 1). For once, the fossil record has spoken in a voice that is more or less unambiguous. ■

Henry Gee is a senior editor of *Nature*.

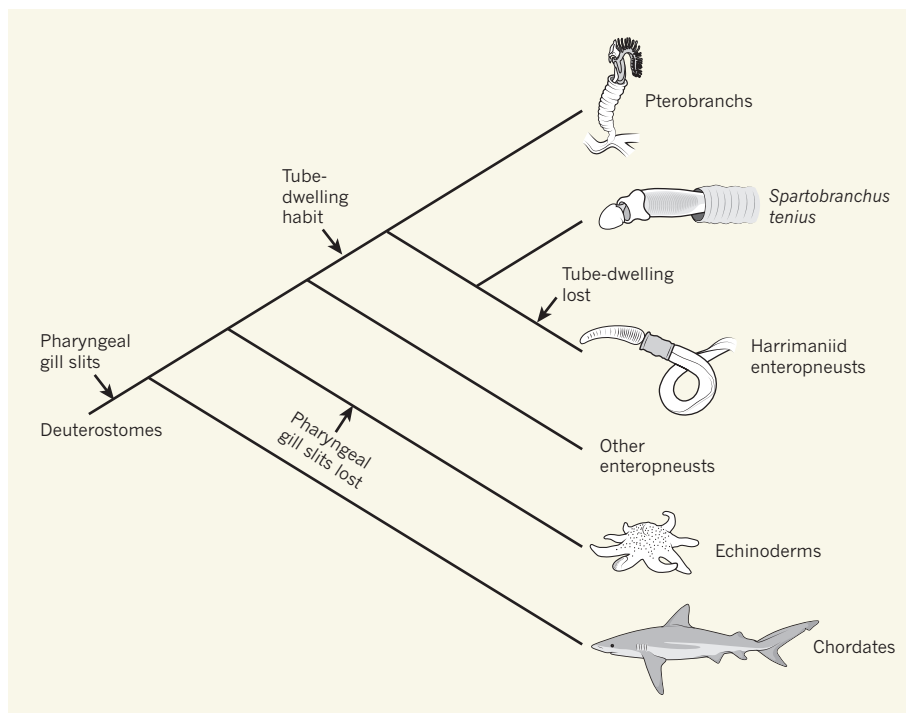


Figure 1 | The habits of our ancestors. This simplified phylogeny shows the evolutionary branching of clades within the deuterostomes, and the stages at which pharyngeal gill slits and a tube-dwelling habit were gained and lost. Caron and colleagues¹ describe enteropneust (acorn worm) fossils that are very similar to modern harrimaniid enteropneusts, but that are associated with tubes, an unusual habit for this otherwise free-living group. The authors name the fossils *Spartobranchus tenuis*. Their discovery suggests that the ancestor of pterobranchs and *S. tenuis* adopted a tube-dwelling habit before pterobranchs became the sessile, colony-dwelling animals that they are today, and that the ancestor of deuterostomes was probably a worm-like, motile animal.

1. Caron, J.-B., Conway Morris, S. & Cameron, C. B. *Nature* **495**, 503–506 (2013).
2. Hall, B. K. *J. Exp. Zool.* **304**, 1–17 (2005).
3. Halanach, K. M. *Mol. Phylogenet. Evol.* **4**, 72–76 (1995).
4. Metchnikoff, V. *E. Zool. Anz.* **4**, 139–157 (1881).
5. Jefferies, R. P. S. *The Ancestry of the Vertebrates* (British Museum (Natural History), 1986).
6. Garstang, W. *Zool. Anz.* **17**, 122–125 (1894).
7. Garstang, W. *Q. J. Microsc. Sci.* **72**, 51–187 (1928).
8. Cameron, C. B., Garey, J. R. & Swalla, B. J. *Proc. Natl Acad. Sci. USA* **97**, 4469–4474 (2000).

X-ray analysis on the nanogram to microgram scale using porous complexes

Yasuhide Inokuma¹, Shota Yoshioka¹, Junko Ariyoshi¹, Tatsuhiko Arai¹, Yuki Hitora², Kentaro Takada², Shigeki Matsunaga², Kari Rissanen³ & Makoto Fujita¹

X-ray single-crystal diffraction (SCD) analysis has the intrinsic limitation that the target molecules must be obtained as single crystals. Here we report a protocol for SCD analysis that does not require the crystallization of the sample. In our method, tiny crystals of porous complexes are soaked in a solution of the target, such that the complexes can absorb the target molecules. Crystallographic analysis clearly determines the absorbed guest structures along with the host frameworks. Because the SCD analysis is carried out on only one tiny crystal of the complex, the required sample mass is of the nanogram–microgram order. We demonstrate that as little as about 80 nanograms of a sample is enough for the SCD analysis. In combination with high-performance liquid chromatography, our protocol allows the direct characterization of multiple fractions, establishing a prototypical means of liquid chromatography SCD analysis. Furthermore, we unambiguously determined the structure of a scarce marine natural product using only 5 micrograms of the compound.

Chemists need reliable methods to analyse and determine molecular structures. Nuclear magnetic resonance (NMR) and mass spectrometry are indispensable tools in daily chemical research for rapidly analysing molecular structures, but, strictly speaking, they provide only speculative molecular structures that are sometimes assigned incorrectly. However, X-ray SCD provides direct structural information at the atomic level and is recognized as the most reliable structure determination method^{1–3}. Unfortunately, X-ray SCD has some critical limitations. First, the crystallization of samples before measurement can not be automated and usually requires a time-consuming trial-and-error procedure. Second, the method is in principle not applicable to non-crystalline molecules. In this Article, we describe an advance in crystallographic analysis based on a new X-ray analysis protocol that does not require the crystallization of the sample molecules themselves.

Our idea is to use networked porous metal complexes^{4–7} as ‘crystalline sponges’⁸. Owing to the high molecular-recognition ability of the pores, the crystalline sponges can absorb target sample molecules from their solution into the pores, rendering the incoming molecules regularly ordered in the crystal. Accordingly, the molecular structure of the absorbed guest will be displayed, along with the host framework, by the crystallographic analysis of the networked porous complexes. We emphasize that even trace amounts of samples (<0.1 µg) can be analysed by this method because the experiment can be performed with only one tiny crystal (<0.1 mm to a side). In the following discussion, we thus describe the crystallographic analysis of non-crystalline compounds and nanogram–microgram-scale X-ray crystallography based on our method. The great advantage of trace-amount X-ray analysis is particularly emphasized by its application to liquid chromatography SCD analysis (see below), where high-performance liquid chromatography (HPLC) fractions are directly collected by the crystalline sponge and analysed by X-ray crystallography. Furthermore, we successfully determine the structure of a scarce marine natural product, miyakosyne A, including the absolute configuration of its chiral centre, which could not be determined by conventional chemical and spectroscopic methods.

X-ray crystallography of liquid samples

In this study, two networked complexes are used as crystalline sponges: $\{[(\text{Co}(\text{NCS})_2)_3(1)_4] \cdot x(\text{solvent})\}_n$ (**2**; ref. 8) and $\{[(\text{ZnI}_2)_3(1)_2] \cdot x(\text{solvent})\}_n$ (**3**; ref. 9). Both of these can be easily prepared from tris(4-pyridyl)-1,3,5-triazine (**1**) and a metal salt ($\text{Co}(\text{NCS})_2$ and ZnI_2 , respectively). Complex **2** has a three-dimensionally networked cage framework, the unit structure of which is identical to that of a discrete M_6L_4 cage host¹⁰ (**4**; M = ethylenediamine-capped Pd(II) ion, L = ligand **1**). The framework of **3** is completely different from that of **2**, but the void spaces of both **2** and **3** show strong binding properties for incoming guest molecules, similar to the molecular recognition by discrete cage **4** in solution. The as-synthesized complexes **2** and **3** contain solvents in the void, which can be exchanged for other guests by soaking the crystals of **2** and **3** in a guest solution. The crystallographic analysis of guest-exchanged **2** and **3** has been reported^{8,11–14}, but primarily a large excess of solid guests have been examined. Focusing on the crystallographic analysis of non-crystalline guests on a small scale, we first developed a procedure for one-crystal-scale guest uptake with only a drop of liquid guest (Fig. 1a). When a crystal of **2** was brought into contact with a drop of cyclohexanone (**5**) at room temperature (25 °C), rapid guest penetration occurred. After 2 d, the crystal was subjected to an X-ray diffraction experiment. As expected, the crystal data clearly showed the molecular structure of **5** packed within the cavity of the unit cage in a host/guest ratio of 1:4 (Fig. 1b). In a similar manner, complex **3** also enclathrated a liquid guest, isoprene (**6**), which is a highly volatile alkene with a melting point below –145 °C (Fig. 1c). Once absorbed into a crystal of **3**, guest **6** was no longer volatile and its structure was easily determined by X-ray crystallography without any special treatment.

It is worth mentioning that the absorbed guests are rarely observed by X-ray analysis with common porous coordination networks^{4–7,15–18}. Such networks are mostly characterized by their permanent vacuum pores, which rapidly absorb small molecules (normally gases), presumably in a kinetic way. In contrast, the void spaces of as-synthesized crystals of **2** and **3** are filled with highly fluid solvent molecules. Guest

¹Department of Applied Chemistry, Graduate School of Engineering, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. ²Laboratory of Aquatic Natural Products Chemistry, Graduate School of Agricultural and Life Science, The University of Tokyo, Yayoi, Bunkyo-ku, Tokyo, 113-8657, Japan. ³Department of Chemistry, NanoScience Center, University of Jyväskylä, PO Box 35, 40014 Jyväskylä, Finland.

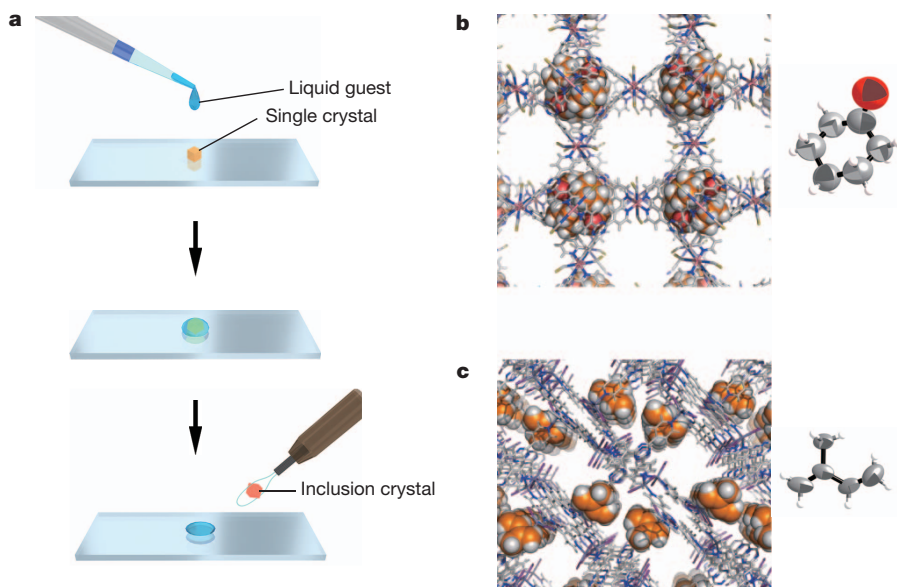


Figure 1 | X-ray crystallographic observation of liquid guest molecules using crystalline sponges. **a**, Schematic outline for the preparation of a guest-included network complex: a single piece of crystal **2** or **3** was treated for 2 d with a drop of a liquid guest and subjected to X-ray data collection. The experiment with isoprene was carried out in a sealed vial to avoid the

molecules can slowly penetrate these ‘wet’ cavities by guest exchange, and are concentrated at the molecular-recognition pockets surrounded by ligand **1**. A characteristic of the strong host–guest interaction in the

evaporation of the guest. **b**, X-ray crystal structure of cyclohexanone (**5**) observed in the cavities of **2**. **c**, Structure of isoprene (**6**) observed in the pores of **3**. In the network structures, liquid guests are shown as space-filling models. Thermal ellipsoids are drawn at the 50% probability level.

crystals of **2** and **3** lies in panel ligand **1**, which attracts various guests onto its electron-deficient π -plane^{19,20}. The entire process takes place under thermodynamic control, similar to host–guest complexation in

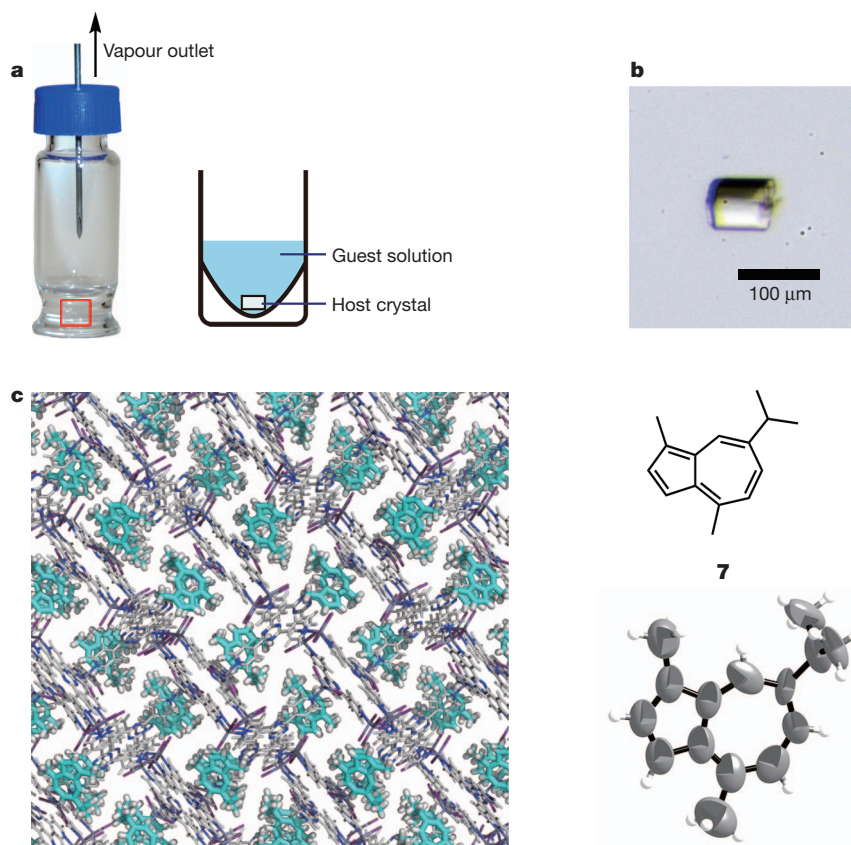


Figure 2 | Nanogram-scale guest inclusion with a crystal of crystalline sponge 3. **a**, Photograph of a microvial including a single crystal of crystalline sponge **3** (solvent, cyclohexane) and a CH_2Cl_2 /cyclohexane (1:9) solution of guaiazulene **7** (500 ng/50 μl). The schematic illustration represents the bottom

of the vial highlighted with a red box in the left picture. **b**, Microgram of the single crystal of **3** placed in a microvial. **c**, Network structure of guaiazulene (**7**) included in crystalline sponge **3** and ORTEP drawing (at the 50% probability) of the guest **7**. (C, grey; H, white).

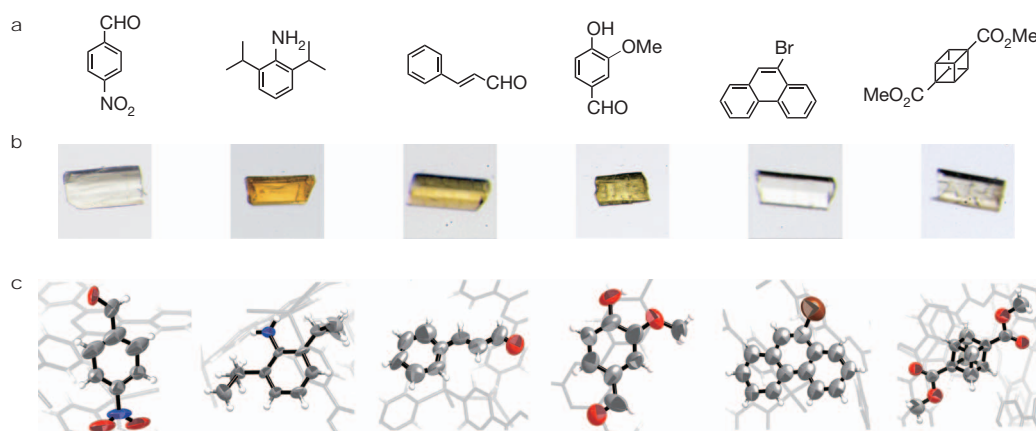


Figure 3 | Crystal structures of a variety of guests determined using a one-crystal-scale inclusion protocol. Guest inclusion and X-ray diffraction measurement were conducted for each compound with 5 μg of sample and one

network crystal **3**. **a**, Chemical structures of guests. Me, methyl. **b**, Inclusion crystals prepared with 5 μg of sample. **c**, ORTEP representation of the crystal structures of the guests in the pores of **3**.

solution. As a result, the geometry of the included guests is well equilibrated and regularly ordered, making the crystallographic analysis of the accommodated guests possible. The equilibration of guest geometry, which is an easily overlooked point in rendering the guests in order, does not normally occur in kinetic guest capture by the permanent pores of coordination networks, so the guests are severely disordered and have rarely been observed by X-ray analysis^{21–25}.

Nanogram–microgram-scale X-ray analysis

A great advantage of this method is particularly clear when it is applied to the structural analysis of trace amounts of samples. Given that the experiment is performed with only one tiny crystal, the required sample amount is estimated to be of the nanogram–microgram order. Because the guest-accessible void of **3** accounts for $\sim 50\%$ of the lattice, one tiny crystal (typically 100 μm to a side; $\sim 1\ \mu\text{g}$) theoretically requires only $\sim 0.5\ \mu\text{g}$ of organic guest (density, $\sim 1\ \text{g cm}^{-3}$), indicating that X-ray crystallographic analysis can be performed on nanogram–microgram scales. Furthermore, if only one of the recognition sites in **3** firmly traps the guest, it is not necessary for the rest of the void space to bind the guest. Thus, the average occupancy of the guest in the crystal can be much lower than $\sim 100\%$, further decreasing the lower limit of the sample amount.

Encouraged by the theoretical estimation, we examined the one-crystal-scale guest inclusion using only 500 ng of guaiazulene (**7**), a natural hydrocarbon. A cube-shaped crystal of **3** (solvent = cyclohexane; $60 \times 60 \times 70\ \mu\text{m}^3$) was placed in a microvial with 45 μl of cyclohexane. A CH_2Cl_2 solution of guest **7** (500 ng, 5 μl) was added to the supernatant and the microvial was capped with a screw cap with a pinhole as an exhaust outlet (Fig. 2). The solvent slowly evaporated through the pinhole in an incubator maintained at 45°C over 2 d. As the volume of supernatant decreased, the crystal of **3**, which was initially colourless, gradually turned dark blue, indicating the concentration of **7** in the

crystal. The crystal was removed from the vial and directly mounted onto an X-ray diffractometer to collect the diffraction data. The crystal structure thus obtained showed **7** trapped in the pores with 60% occupancy (Fig. 2c). Importantly, two methyl and one isopropyl substituent on the azulene core were readily apparent from the observed electron density map.

From the occupancy of **7**, the total amount of the guest actually included in the crystal was estimated to be only 26 ng, which accounts for less than 10% of the guest added to the supernatant. This result motivated us to reduce the guest amount further. Surprisingly, when only 80 ng of guest **7** was used with a crystal of **3** ($80 \times 80 \times 80\ \mu\text{m}^3$), guest **7** was still clearly observed. Considering that the experiment was carried out using a laboratory X-ray machine, we presume that crystallography is possible with synchrotron X-ray experiments even on a mass of $< 10\ \text{ng}$.

For practical reasons, a typical protocol for trace-amount crystallography was established with 5 μg of sample (Methods Summary). Six appropriate samples were selected and one of us (S.Y.) performed the trace-amount crystallography with only $\sim 5\ \mu\text{g}$ of each sample and without any knowledge of the structures. In this blind test, three structures were fully determined from only the diffraction data (Fig. 3). The other three were initially flawed because of atom misassignment, symmetry problems and guest disorder, which are common problems in crystallographic analysis. However, the incorrect structures could be easily corrected using only the mass spectrometric data (molecular weight information). Thus, like common crystallographic analysis, the refined crystal structures should be supported by other methods of structural analysis such as mass and NMR spectrometry. The successful structural determination of these compounds indicates that the scope of guests can be considerably extended to include polycyclic, non-aromatic and non-planar molecules.

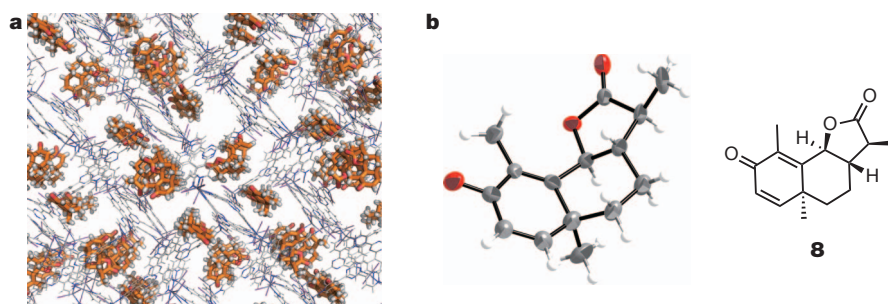


Figure 4 | The crystal structure of a chiral guest, santonin, trapped in a crystalline sponge. **a**, Network structure of the clathrate comprising santonin

(**8**) and complex **3**. **b**, ORTEP drawing (50% probability) of the santonin trapped in the pore, and chemical structure of santonin.

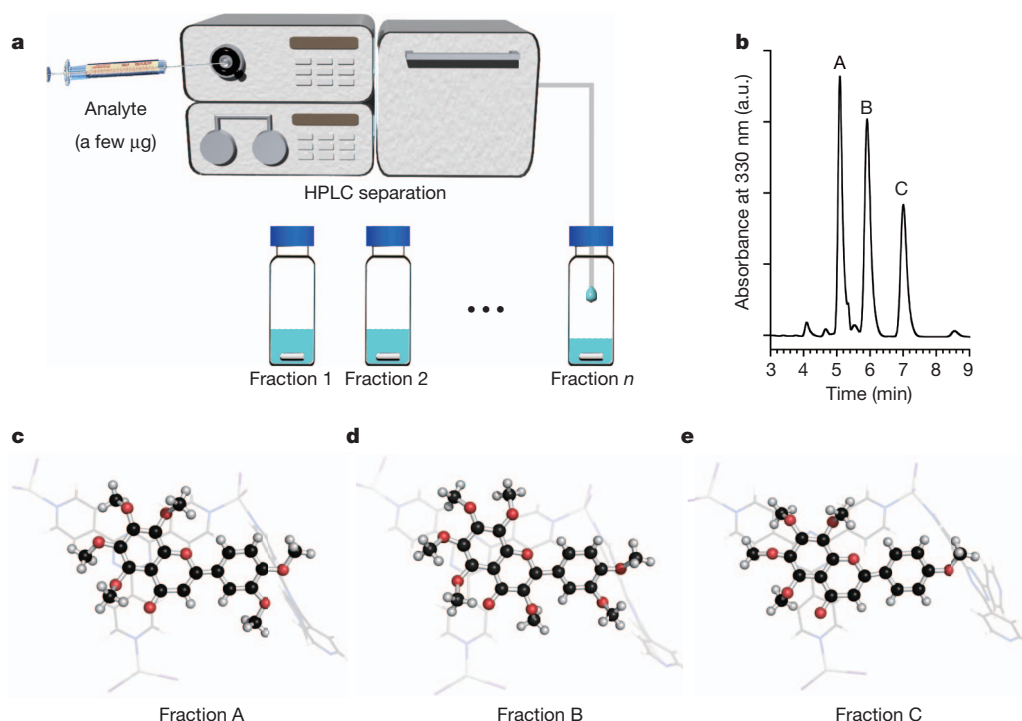


Figure 5 | LC-SCD analysis of natural flavonoids. **a**, Sketch of LC-SCD analysis. Each fraction isolated by analytical HPLC is poured into a microvial that contains a single crystal of crystalline sponge. **b**, HPLC chromatogram of PMFs extracted from the peel of *Citrus unshiu*. a.u., arbitrary units. **c–e**, X-ray

crystal structures of fraction A, nobiletin (5,6,7,8,3',4'-hexamethoxyflavone) (**9**; **c**); fraction B (3,5,6,7,8,3',4'-heptamethoxyflavone) (**10**; **d**); and fraction C, tangeritin (5,6,7,8,4'-pentamethoxyflavone) (**11**; **e**) obtained by LC-SCD analysis. (C, black; O, red; H, grey).

Absolute structure determination

The determination of the absolute configuration of chiral compounds is one of the most difficult analyses of molecular structures^{26–28}. NMR and other spectroscopic methods can determine in principle only relative stereochemistry. X-ray crystallography is the only method that can determine the absolute configuration of chiral molecules, on the basis of the anomalous scattering effects of heavy atoms^{29,30}. Surprisingly, we were able to determine the absolute structure of a trace amount (5 µg) of a chiral compound because the heavy atoms (Zn and I) in the host framework (**3**) showed clear anomalous scattering. As-synthesized coordination network **3** has an achiral space group ($C2/c$), but when treated with 5 µg of santonin (**8**), an anthelmintic drug bearing four chiral centres³¹, the space group changed to chiral $P2_1$ owing to the enclathration of the chiral guest in the pores. The refined absolute structure was validated by the Flack parameter (0.092(18); parenthetical error, s.d.) and the stereochemistry of **8** was unambiguously determined (Fig. 4). It is noteworthy that, in contrast to common absolute structure determinations^{32–34}, the chemical introduction of heavy atoms on the guest skeleton is unnecessary because the host framework contains heavy atoms and shows enhanced anomalous scattering effects. We believe that our method is the most practical and reliable method for the determination of the absolute structure of chiral compounds.

LC-SCD analysis

The requisite guest amount in our protocol matches the separation scale of analytical HPLC (a few micrograms). This prompted us to propose a combination of liquid chromatography and SCD, namely LC-SCD analysis. In this analysis, a trace amount of a sample mixture is separated by HPLC and each fraction is directly treated with crystalline sponge **3**. Subsequently the molecular structure of each fraction is analysed by X-ray crystallography. As a proof of concept, we considered the structural determination of a series of polymethoxyflavones (PMFs) that are trace components in orange peel^{35,36}. A crude mixture of PMFs (~30 µg) extracted from a piece of air-dried orange

peel, *Citrus unshiu*, was separated by analytical HPLC to give three main components, A, B and C (Fig. 5), that weighted 6, 7 and 5 µg, respectively. Each fraction was then directly subjected to the guest inclusion conditions with one single crystal of **3**. The X-ray crystal structures of the three samples all showed a flavone ring trapped at a similar position in the host pores. On the basis of the X-ray structures, components A, B and C were identified as PMFs **9**, **10**, and **11**, respectively. These structures are in good accordance with their mass data (Supplementary Information). It is worth noting that, despite the PMFs' small, simple structures, the determination of their substitution positions is unexpectedly difficult even with the use of NMR spectroscopy. Considering the applicable range of our X-ray protocol, the LC-SCD analysis will be a powerful tool for the rapid characterization of multiple components with much higher structural reliability than liquid chromatography mass spectrometry and LC-NMR techniques.

Complete stereochemistry of miyakosyne A

Using our method, we took on the considerable challenge of determining the structure of miyakosyne A, a scarce natural product. Miyakosyne A (**12**), a marine natural product recently isolated from a marine sponge *Petrosia* sp., has three chiral centres on its main alkyl chain³⁷ (Fig. 6a). Although the absolute configurations at C3 and C26 have been previously determined to be 3*R* and 26*R*, respectively, conventional spectroscopic and chemical methods are no longer effective for the determination of the absolute configuration at C14 because the difference between the two long alkyl groups on C14 is only one methylene unit. X-ray crystallography is expected to be the only method that can determine the C14 configuration, but preparation of a single crystal of **12** from only a limited amount is an extremely challenging task. Hence, we applied our method to the full characterization of miyakosyne A to determine the absolute configuration at C14.

A tiny crystal of **3** (approximately $100 \times 100 \times 100 \mu\text{m}^3$) was treated with 5 µg of miyakosyne A (**12**) and subjected to a diffraction

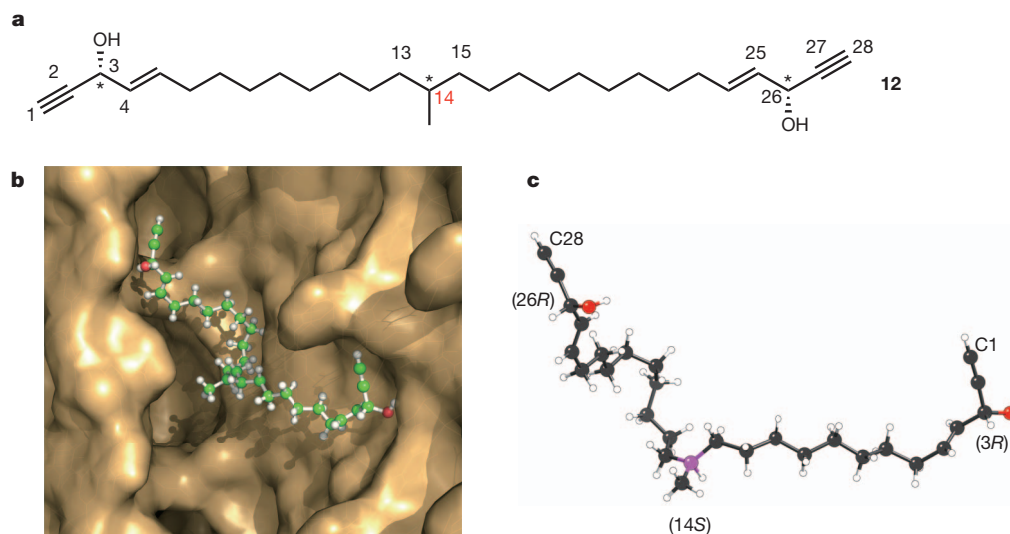


Figure 6 | Structural determination of miyakosyne A. **a**, Chemical structure of miyakosyne A (**12**). Numbers indicate carbon atoms; asterisks indicate asymmetric carbon atoms. **b**, Orientation of miyakosyne A enclathrated in the pore of crystalline sponge **3**. The host framework **3** is drawn as a brown surface,

and the guest is represented as a ball-and-stick model. (C, green; O, red; H, white) **c**, The absolute structure of miyakosyne A determined by relative configuration to C3 and C26. The chiral centre C14 is drawn in magenta.

study. The crystal structure was successfully solved in the chiral C2 space group and revealed all the stereochemistry of **12**. On the basis of the known 3R and 26R configurations, the C14 configuration of **12** was determined to be S (Fig. 6b, c). Because the occupancy of the solved guest **12** was ~50% due to disordered guests and remaining solvents in the pores, we carefully checked the validation of the crystal structure. We found that the observed electron density map ($2F_o - F_c$) can be well superimposed on the refined structure of (3R,14S,26R)-**12**, and that when least-squares refinement was performed with an incorrectly modelled 14R form without any restraints around C14, the refinement did not converge to this configuration but converged finally to the 14S form. This result is a striking example of the structural determination of a scarce natural product and shows that our method can be expanded in scope to cover the structural elucidation of even large and conformationally flexible compounds. A variety of different types of compound can be quickly and readily analysed in a similar way.

Conclusion

Our method solves the real and intrinsic problems of X-ray crystallography and transforms it into a rapid and convenient method for the analysis of molecular structures using only a trace amount of sample. The following features are worthy of special attention. First, the crystallization step, which is the bottleneck of the X-ray analysis protocol, becomes unnecessary. Therefore, the crystallographic study of molecular structures is drastically accelerated and is now applicable to the analysis of liquid or even volatile compounds. Second, crystallographic analysis can be preformed on the nanogram–microgram scale. Thus, in terms of sensitivity, X-ray analysis overwhelmingly dominates NMR analysis and is even comparable to mass spectrometry. The combination of HPLC and single-crystal X-ray analysis (LC–SCD analysis) has been achieved. Third, the determination of the absolute configuration of chiral molecules can be easily carried out without any chemical modification of the sample molecules.

The slow diffusion of guests into the solvent-filled voids of the crystalline sponges is a particularly important process that renders the guests thermodynamically well equilibrated at the specific molecular-recognition sites of the hosts. Accordingly, the guests are regularly ordered and made observable by crystallographic analysis. This is not always the case with common porous coordination networks, in which the permanent pores absorb guests in presumably a

kinetic way. Because the crystalline sponges used here are easily prepared from commercially available components, the nanogram–microgram-scale crystallographic analysis of non-crystalline molecules is already practicable. We expect that our method will be applicable to microanalysis in natural product chemistry, food and perfume science, drug discovery, forensic science, and, most importantly, the daily research of synthetic chemists^{38–40}. We believe that many natural and synthetic compounds that chemists have almost given up hope of analysing crystallographically will be easily and precisely characterized by this method.

METHODS SUMMARY

A cube-shaped crystal of **3** (approximately $100 \times 100 \times 100 \mu\text{m}^3$) was placed in the bottom of a microvial with 45 μl of cyclohexane. A solution of guest compound (5 μg) in dichloromethane (5 μl) was added to the supernatant, and the microvial was capped with a screw cap pierced with a hollow needle (internal diameter, 0.8 mm). The solvent of the supernatant was slowly evaporated in an incubator maintained at 45 °C over 2 d at a rate of $\sim 1 \mu\text{l h}^{-1}$. The guest inclusion crystal was directly mounted onto a single-crystal X-ray diffractometer with a cryoloop and the diffraction data were collected. The crystal structures were solved and refined by using the SHELX 97 program. Further details and other methods are available in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper.

Received 13 November 2012; accepted 5 February 2013.

- Ooi, L. *Principles of X-Ray Crystallography* (Oxford Univ. Press, 2010).
- Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
- Ohashi, Y. in *Models, Mysteries and Magic of Molecules* (eds Boeyens, J. C. A. & Ogilvie, J. F.) 109–113 (Springer, 2008).
- Batten, S. R. & Robson, R. Interpenetrating nets: ordered, periodic entanglement. *Angew. Chem. Int. Ed.* **37**, 1460–1494 (1998).
- Kitagawa, S., Kitaura, R. & Noro, S. Functional porous coordination polymers. *Angew. Chem. Int. Ed.* **43**, 2334–2375 (2004).
- Yaghi, O. M. *et al.* Reticular synthesis and the design of new materials. *Nature* **423**, 705–714 (2003).
- Fujita, M., Kwon, Y. J., Washizu, S. & Ogura, K. Preparation, clathration ability and catalysis of a two-dimensional square network material composed of cadmium(II) and 4,4'-bipyridine. *J. Am. Chem. Soc.* **116**, 1151–1152 (1994).
- Inokuma, Y., Arai, T. & Fujita, M. Networked molecular cages as crystalline sponges for fullerenes and other guests. *Nature Chem.* **2**, 780–783 (2010).
- Biradha, K. & Fujita, M. A springlike 3D-coordination network that shrinks or swells in a crystal-to-crystal manner upon guest removal or readsorption. *Angew. Chem. Int. Ed.* **41**, 3392–3395 (2002).
- Fujita, M. *et al.* Self-assembly of ten molecules into nanometre-sized organic host frameworks. *Nature* **378**, 469–471 (1995).

11. Inokuma, Y., Kojima, N., Arai, T. & Fujita, M. Bimolecular reaction via the successive introduction of two substrates into the crystals of networked molecular cages. *J. Am. Chem. Soc.* **133**, 19691–19693 (2011).
12. Ohmori, O., Kawano, M. & Fujita, M. Crystal-to-crystal guest exchange of large organic molecules within a 3D coordination network. *J. Am. Chem. Soc.* **126**, 16292–16293 (2004).
13. Haneda, T., Kawano, M., Kojima, T. & Fujita, M. Thermo-to-photo-switching of the chromic behavior of salicylideneanilines by inclusion in a porous coordination network. *Angew. Chem. Int. Ed.* **46**, 6643–6645 (2007).
14. Ohara, K., Kawano, M., Inokuma, Y. & Fujita, M. A porous coordination network catalyzes an olefin isomerization reaction in the pore. *J. Am. Chem. Soc.* **132**, 30–31 (2010).
15. Férey, G. Hybrid porous solids: past, present, future. *Chem. Soc. Rev.* **37**, 191–214 (2008).
16. Li, J.-R., Kuppler, R. J. & Zhou, H.-C. Selective gas adsorption and separation in metal-organic frameworks. *Chem. Soc. Rev.* **38**, 1477–1504 (2009).
17. Chen, B., Xiang, S. & Qian, G. Metal-organic frameworks with functional pores for recognition of small molecules. *Acc. Chem. Res.* **43**, 1115–1124 (2010).
18. Kondo, M. *et al.* Three-dimensional framework with channeling cavities for small molecules: $\{[M_2(4,4'\text{-bpy})_3(\text{NO}_3)_4] \cdot x\text{H}_2\text{O}\}_n$ ($M = \text{Co}, \text{Ni}, \text{Zn}$). *Angew. Chem. Int. Edn Engl.* **36**, 1725–1727 (1997).
19. Yoshizawa, M., Klosterman, J. K. & Fujita, M. Functional molecular flasks: new properties and reactions within discrete, self-assembled hosts. *Angew. Chem. Int. Ed.* **48**, 3418–3438 (2009).
20. Inokuma, Y., Kawano, M. & Fujita, M. Crystalline molecular flasks. *Nature Chem.* **3**, 349–358 (2011).
21. Li, Q. W. *et al.* Docking in metal-organic frameworks. *Science* **325**, 855–859 (2009).
22. Kim, H., Chun, H., Kim, G.-H., Lee, H.-S. & Kim, K. Vapor phase inclusion of ferrocene and its derivative in a microporous metal-organic porous material and its structural characterization by single crystal X-ray diffraction. *Chem. Commun.* 2759–2761 (2006).
23. Halder, G. J. & Kepert, C. J. In situ single-crystal X-ray diffraction studies of desorption and sorption in a flexible nanoporous molecular framework material. *J. Am. Chem. Soc.* **127**, 7891–7900 (2005).
24. Kawano, M. & Fujita, M. Direct observation of crystalline-state guest exchange in coordination networks. *Coord. Chem. Rev.* **251**, 2592–2605 (2007).
25. Kitaura, R. *et al.* Formation of a one-dimensional array of oxygen in a microporous metal-organic solid. *Science* **298**, 2358–2361 (2002).
26. Cahn, R. S., Ingold, C. & Prelog, V. Specification of molecular chirality. *Angew. Chem. Int. Edn Engl.* **5**, 385–415 (1966).
27. Seco, J. M., Quiñoa, E. & Riguera, R. The assignment of absolute configuration by NMR. *Chem. Rev.* **104**, 17–118 (2004).
28. Freedman, T. B., Cao, X., Dukor, R. K. & Nafie, L. A. Absolute configuration determination of chiral molecules in the solution state using vibrational circular dichroism. *Chirality* **15**, 743–758 (2003).
29. Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. Determination of the absolute configuration of optically active compounds by means of X-rays. *Nature* **168**, 271–272 (1951).
30. Flack, H. D. & Bernardinelli, G. Absolute structure and absolute configuration. *Acta Crystallogr. A* **55**, 908–915 (1999).
31. Corey, E. J. The stereochemistry of santonin, β -santonin, and artemisin. *J. Am. Chem. Soc.* **77**, 1044–1045 (1955).
32. Deschamps, J. R. X-ray crystallography of chemical compounds. *Life Sci.* **86**, 585–589 (2010).
33. Takayanagi, H., Sudou, M. & Ogura, H. Crystal structure of $1\alpha,2\beta$ -dibromo- $1,2$ -dihydro- α -santonin. *Anal. Sci.* **7**, 183–184 (1991).
34. Inayama, S. *et al.* Unusual bromination of tetrahydro-(γ)- α -santonins and new santonin isomers: X-ray crystal and molecular structure of $2\beta,14$ -dibromo- $4\alpha,5\beta,6\beta,11\beta$ H-tetrahydrosantonin. *J. Chem. Soc. Chem. Commun.* 495–496 (1980).
35. Green, C. O., Wheatley, A. O., Osagie, A. U., Morrison, E. Y. S. A. & Asemota, H. N. Determination of polymethoxylated flavones in peels of selected Jamaican and Mexican citrus (*Citrus* spp.) cultivars by high-performance liquid chromatography. *Biomed. Chromatogr.* **21**, 48–54 (2007).
36. Han, S. *et al.* Isolation and identification of polymethoxyflavones from the hybrid *Citrus*, Hallabong. *J. Agric. Food Chem.* **58**, 9488–9491 (2010).
37. Hitora, Y., Takada, K., Okada, S. & Matsunaga, S. Miyakosynes A–F, cytotoxic methyl branched acetylenes from a marine sponge *Petrosia* sp. *Tetrahedron* **67**, 4530–4534 (2011).
38. Sampietro, D. A., Catalan, C. A. N. & Vattuone, M. A. *Isolation, Identification, and Characterization of Allelochemicals/Natural Products* (Science Publ., 2009).
39. Croue, J.-P., Korshin, G. V. & Benjamin, M. M. *Characterization of Natural Organic Matter in Drinking Water* 73–374 (Am. Water Works Assoc., 1999).
40. Ahuja, S. & Alsante, K. *Handbook of Isolation and Characterization of Impurities in Pharmaceuticals* (Academic, 2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by Grants-in-Aid for Specially Promoted Research (24000009) and Young Scientists (B) (23750146), and by the CREST project of the Japan Science and Technology Agency. The experiment involving X-ray crystallography with 80 ng of guest molecules was performed using VariMax optics with a RAPID image plate detector system, courtesy of Rigaku Corporation. We thank M. Yamasaki and H. Sato for support for X-ray measurements.

Author Contributions Y.I. and M.F. designed the project, analysed results and wrote the manuscript. S.Y., J.A. and T.A. performed the experimental work and crystallographic analysis. Y.H., S.M. and K.T. selected and provided a natural product sample for analysis. K.R. confirmed the validity of the X-ray crystallographic analysis of all data.

Author Information The X-ray crystallographic coordinates for structures reported in this paper have been deposited at the Cambridge Crystallographic Data Centre, under deposition numbers CCDC 910380, 910381, 910382, 910383, 910384, 910385, 910386, 910387, 910388, 910389, 910390, 910391, 910392, 910393 and 910394. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk/data_request/cif). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.F. (mfujita@appchem.t.u-tokyo.ac.jp).

METHODS

Chemicals. Solvents and reagents were purchased from TCI, WAKO Pure Chemical Industries and Sigma-Aldrich, and were used without any further purification except where noted.

Glassware. Microvials for this research were purchased from Waters (Deactivated Clear Glass 12 × 32 mm Screw Neck Max Recovery Vial; 1.5 ml). A syringe needle, Terumo NN-2116R (internal diameter, 0.80 mm), was used as an exhaust outlet for the cap of the microvial. An incubator (Fine FF-12) was used to maintain the temperature during guest inclusion.

Equipment. HPLC separation of microgram-scale natural products was performed on a Jasco UV-970 spectrometer equipped with a Jasco PU-980 pump and an ODS column Develosil ODS-5, (Nomura Chemical Co.). Elemental analysis was performed on a CHN/O/S elemental analyser (CE-440, Exeter Analytical). Mass spectra were recorded on a Bruker maXis spectrometer. Single-crystal X-ray diffraction data were collected on a Bruker APEX-II CCD diffractometer equipped with a focusing mirror (MoK α radiation wavelength, 0.71073 Å) and an N $_2$ generator (Japan Thermal Engineering Co.) or a Rigaku VariMax with RAPID (CuK α radiation wavelength, 1.54187 Å). Microscopic infra-red spectra were recorded on a Varian DIGILAB Scimitar instrument. For X-ray diffraction and microscopic infrared measurements, fluorolube was used as a protectant for the single crystals.

X-ray structural analysis. The structures were solved by direct methods and refined by full-matrix least-squares calculations on F^2 (the squared structure

factor), using the SHELX97 program. Residual electron densities in the solvent-accessible void due to disordered solvent molecules were treated with the PLATON/SQUEEZE program. The quite large ($3e-9e\text{ Å}^{-3}$) residual electron density peaks causing checkCIF B level alerts in SQUEEZE are all located near the iodine atoms in the corresponding structures and could not be modelled with reasonably disordered atoms.

Plant material. *Citrus unshiu* was cultivated in Nishi-Uwajima, Ehime prefecture, Japan in 2011. The peel of *C. unshiu* was air-dried at room temperature over 1 week and ground into a fine powder with a lab mixer.

Inclusion of cyclohexanone (5) into a crystal of 2. On a glass plate, a crystal of **2** ($280 \times 280 \times 200\text{ }\mu\text{m}^3$) was soaked in a drop of cyclohexanone (**5**) (5 μl) at room temperature. After standing for 2 d, the crystal was picked using a protectant and mounted onto the X-ray diffractometer. After collection of the diffraction data, the same crystal was subsequently analysed by microscopic Fourier transform infrared spectroscopy.

Inclusion of isoprene (6) into a crystal of 3. In a microvial, a crystal of **3** ($80 \times 50 \times 50\text{ }\mu\text{m}^3$) was soaked in a drop of isoprene (**6**) (5 μl). Then the vial was sealed with a screw cap and allowed to stand for 2 d at room temperature. The crystal was then analysed as described above.

Nanogram-scale inclusion of guaiazulene (7). To a microvial containing a single crystal of **3** (approximately $80 \times 80 \times 80\text{ }\mu\text{m}^3$) and cyclohexane (45 μl), we added a dichloromethane solution of guaiazulene (**7**) (80 ng/5 μl or 500 ng/5 μl). The crystal was heated at 45 °C for 2 d, as described in the general procedure.

Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS

Hong Joo Kim^{1*}, Nam Chul Kim^{1*}, Yong-Dong Wang^{2*}, Emily A. Scarborough^{3*}, Jennifer Moore^{1*}, Zamia Diaz^{3*}, Kyle S. MacLea⁴, Brian Freibaum¹, Songqing Li¹, Amandine Molliex¹, Anderson P. Kanagaraj¹, Robert Carter⁵, Kevin B. Boylan⁶, Aleksandra M. Wojtas⁶, Rosa Rademakers⁶, Jack L. Pinkus⁷, Steven A. Greenberg⁷, John Q. Trojanowski⁸, Bryan J. Traynor⁹, Bradley N. Smith⁸, Simon Topp¹⁰, Athina-Soragia Gkazi¹⁰, Jack Miller¹⁰, Christopher E. Shaw¹⁰, Michael Kottlors¹¹, Janbernd Kirschner¹¹, Alan Pestronk¹², Yun R. Li¹³, Alice Flynn Ford³, Aaron D. Gitler¹⁴, Michael Benatar¹⁵, Oliver D. King¹⁶, Virginia E. Kimonis¹⁷, Eric D. Ross⁴, Conrad C. Weihl¹², James Shorter³ & J. Paul Taylor¹

Algorithms designed to identify canonical yeast prions predict that around 250 human proteins, including several RNA-binding proteins associated with neurodegenerative disease, harbour a distinctive prion-like domain (PrLD) enriched in uncharged polar amino acids and glycine. PrLDs in RNA-binding proteins are essential for the assembly of ribonucleoprotein granules. However, the interplay between human PrLD function and disease is not understood. Here we define pathogenic mutations in PrLDs of heterogeneous nuclear ribonucleoproteins (hnRNPs) A2B1 and A1 in families with inherited degeneration affecting muscle, brain, motor neuron and bone, and in one case of familial amyotrophic lateral sclerosis. Wild-type hnRNPA2 (the most abundant isoform of hnRNPA2B1) and hnRNPA1 show an intrinsic tendency to assemble into self-seeding fibrils, which is exacerbated by the disease mutations. Indeed, the pathogenic mutations strengthen a 'steric zipper' motif in the PrLD, which accelerates the formation of self-seeding fibrils that cross-seed polymerization of wild-type hnRNP. Notably, the disease mutations promote excess incorporation of hnRNPA2 and hnRNPA1 into stress granules and drive the formation of cytoplasmic inclusions in animal models that recapitulate the human pathology. Thus, dysregulated polymerization caused by a potent mutant steric zipper motif in a PrLD can initiate degenerative disease. Related proteins with PrLDs should therefore be considered candidates for initiating and perhaps propagating proteinopathies of muscle, brain, motor neuron and bone.

Elucidating the genetic basis of rare, inherited diseases can provide valuable insights to the molecular pathogenesis of common diseases. Inclusion body myopathy (IBM) associated with Paget's disease of the bone (PDB), fronto-temporal dementia (FTD) and amyotrophic lateral sclerosis (ALS)—sometimes called IBMPFD/ALS—is a rare disorder characterized by progressive degeneration of muscle, brain, motor neurons and bone accompanied by prominent TDP-43 (also called TARDBP) pathology¹. Patients with this rare, inherited syndrome experience features of IBM, FTD, ALS or PDB indistinguishable from familial and sporadic cases of these disorders, and the disease may manifest in several tissues in the same patient^{1,2}. Recently the name multisystem proteinopathy (MSP) has been adopted to reflect the expanding phenotype and prominent proteinaceous pathology that characterizes this syndrome. Some, but not all, cases of MSP are caused by mutations in the *VCP* gene³, which encodes the AAA+ ATPase VCP, a ubiquitin-dependent segregase.

The discovery that *VCP* mutations cause MSP led to the subsequent discovery of pathogenic *VCP* mutations in more common diseases such as sporadic or familial forms of ALS², FTD⁴, IBM⁵ and PDB⁶. These rare MSP families represent a unique opportunity to identify fundamental molecular defects shared among age-related diseases; thus it is highly desirable to identify additional genetic mutations responsible for this syndrome.

A pathogenic mutation in hnRNPA2B1 in MSP

We identified a family (family 1) with dominantly inherited degeneration affecting muscle, bone, brain and motor neurons that was clinically indistinguishable from previous families we have seen with *VCP*-related MSP (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1). Sequencing the entire *VCP* gene (including introns and exons) in affected patients revealed no synonymous or nonsynonymous variants. Genetic analysis of this family by exome sequencing and linkage analysis

¹Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38120, USA. ²Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, Tennessee 38120, USA. ³Department of Biochemistry and Biophysics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

⁴Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523, USA. ⁵Department of Computational Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38120, USA. ⁶Department of Neuroscience, Mayo Clinic, Jacksonville, Florida 32224, USA. ⁷Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Department of Pathology and Laboratory Medicine, Institute on Aging and Center for Neurodegenerative Disease Research, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁹Neuromuscular Diseases Research Group, Laboratory of Neurogenetics, Porter Neuroscience Building, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁰King's College London Centre for Neurodegeneration Research, Department of Clinical Neuroscience, Institute of Psychiatry, London SE5 8AF, UK.

¹¹Division of Neuropediatrics and Muscle Disorders, University Children's Hospital Freiburg, 79106 Freiburg, Germany. ¹²Department of Neurology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. ¹³Medical Scientist Training Program, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹⁴Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ¹⁵Neurology Department, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. ¹⁶Boston Biomedical Research Institute, Watertown, Massachusetts 02472, USA. ¹⁷Department of Pediatrics, Division of Genetics and Metabolism, University of California-Irvine, 2501 Hewitt Hall, Irvine, California 92696, USA.

*These authors contributed equally to this work.

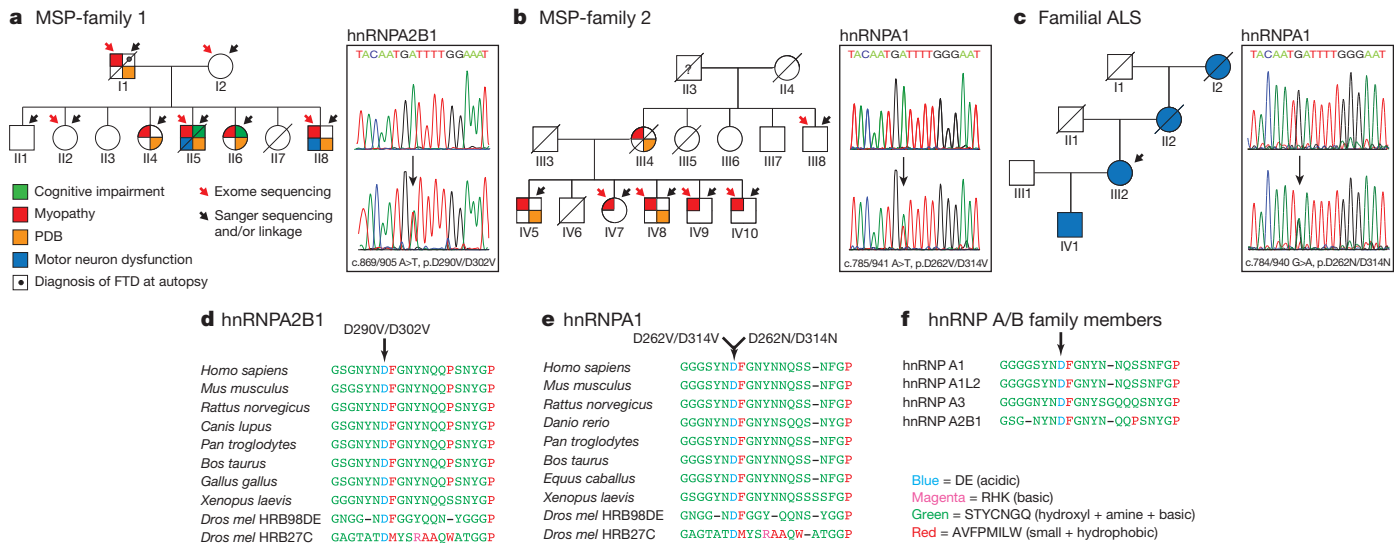


Figure 1 | Identification of previously unknown disease mutations in MSP and ALS. **a**, Family 1 pedigree indicating individuals affected by dementia, myopathy, PDB and ALS. The causative mutation was p.D290V/D302V in hnRNA2B1. Roman numerals denote generation and Arabic numbers denote family member within a generation. **b**, Family 2 pedigree indicating individuals affected by myopathy and PDB. The causative mutation was p.D262V/D314V in hnRNA1. **c**, The pedigree of a family with ALS. The causative mutation was p.D262N/D314N in hnRNA1. **d, e**, Sequence alignment of hnRNA2B1 (**d**) and hnRNA1 (**e**) orthologues showing evolutionary conservation of the

in parallel (Supplementary Fig. 1) identified a single, previously unknown variant (c.869/905A>T, p.D290V/D302V) that co-segregates with disease and alters the gene encoding hnRNA2B1, a ubiquitously expressed RNA-binding protein (Fig. 1a). hnRNA2B1 is expressed as two alternatively spliced isoforms: A2 and B1. The shorter hnRNA2, which lacks 12 amino acids in the amino-terminal region, is the main isoform, accounting for ~90% of the protein in most tissues. The identified mutation substitutes a valine residue in the place of an evolutionarily conserved aspartate residue (Fig. 1d) that is centred in a motif conserved in multiple human paralogs of the hnRNP A/B family (Fig. 1f and Supplementary Fig. 3).

Pathogenic mutations in hnRNA1 in MSP and ALS

Additional validation of the pathogenicity of the hnRNA2B1 mutation came from the analysis of family 2. The clinical features of this family with VCP-negative MSP have been previously described⁷. To identify the culprit mutation we took the same strategy as for family 1 (Supplementary Fig. 2 and Supplementary Table 2). This analysis identified five previously unknown single nucleotide variants and one indel that co-segregated with the disease. Of these, a mutation in hnRNA1 (c.785/941A>T, p.D262V/D314V) stood out because it was the only variant associated with a conserved residue, it was predicted to be deleterious, and because hnRNA1 is highly expressed in affected tissues (Fig. 1b and Supplementary Fig. 2). Moreover, this mutation in hnRNA1 was identical to that found in hnRNA2B1 in family 1, creating a substitution of a valine residue in the place of a highly conserved aspartate residue that is centred in a motif conserved in multiple human paralogs of the hnRNP A/B family (Fig. 1e, f and Supplementary Fig. 3). Subsequent to the identification of these MSP-causing mutations we screened the exomes of 212 familial ALS cases for sequence variants affecting hnRNA2B1 or hnRNA1. In one dominantly inherited case in which known ALS genes were formally excluded, we identified a mutation (c.784/940G>A; p.D262N/D314N) affecting the identical, conserved aspartate residue in hnRNA1 (Fig. 1c, e, f).

The identification of disease mutations in hnRNA2B1 and hnRNA1 was intriguing for three reasons. First, hnRNA2B1 and hnRNA1 interact directly with TDP-43 and function cooperatively

to regulate RNA metabolism⁸. Second, an unbiased genetic screen previously identified fly homologues of TDP-43, hnRNA2B1 and hnRNA1 as suppressors of VCP-related degeneration in a *Drosophila* model of MSP⁹ (Supplementary Fig. 3). Last, hnRNA2B1 has been previously implicated in neurodegenerative disease. Specifically, hnRNA2B1 is sequestered in RNA foci in the fragile-X-associated tremor ataxia syndrome (FXTAS)¹⁰, binds the expanded riboCGG repeats that underlie this disease^{11,12}, and is a genetic modifier in a *Drosophila* model of FXTAS^{11,12}.

hnRNA2B1 and hnRNA1 pathology in MSP

Muscle biopsies from patients II5 (family 1) and IV9 (family 2) showed atrophic fibres, central nuclei and rimmed vacuoles characteristic of IBM (Supplementary Fig. 4a–c). Whereas in normal muscle hnRNA2B1 and hnRNA1 are exclusively nuclear (Fig. 2a, e and Supplementary Fig. 4d), analysis of muscle tissue from patient II5 showed that hnRNA2B1 cleared from many nuclei and accumulated in cytoplasmic inclusions in ~10% of fibres (Fig. 2b and Supplementary Fig. 4e). Muscle from this patient also showed TDP-43 pathology consisting of nuclear clearance and cytoplasmic inclusions, consistent with previous observations in VCP-related and sporadic IBM (Fig. 2j and Supplementary Fig. 4p)¹³. Interestingly, hnRNA2B1 pathology was also observed in VCP-related and sporadic IBM (Fig. 2c, d and Supplementary Fig. 4h, i).

Analysis of muscle from patient IV9 also revealed nuclear clearance and cytoplasmic inclusions of hnRNA1 in ~10% of fibres (Fig. 2f and Supplementary Fig. 4j). hnRNA2B1 and TDP-43 pathology were also observed in the muscle biopsy from patient IV9 (Fig. 2h, k and Supplementary Fig. 4f, g). Moreover, hnRNA1 pathology was also observed in VCP-related and sporadic IBM (Fig. 2g and data not shown). Therefore, regardless of aetiology we have found hnRNA2B1, hnRNA1 and TDP-43 pathology in sporadic and familial IBM. Pathological redistribution of hnRNA2B1, hnRNA1 and TDP-43 was not observed in other types of muscle disease (Supplementary Fig. 4q). FUS (also known as TLS)-associated pathology was not observed in these cases, and thus pathological redistribution

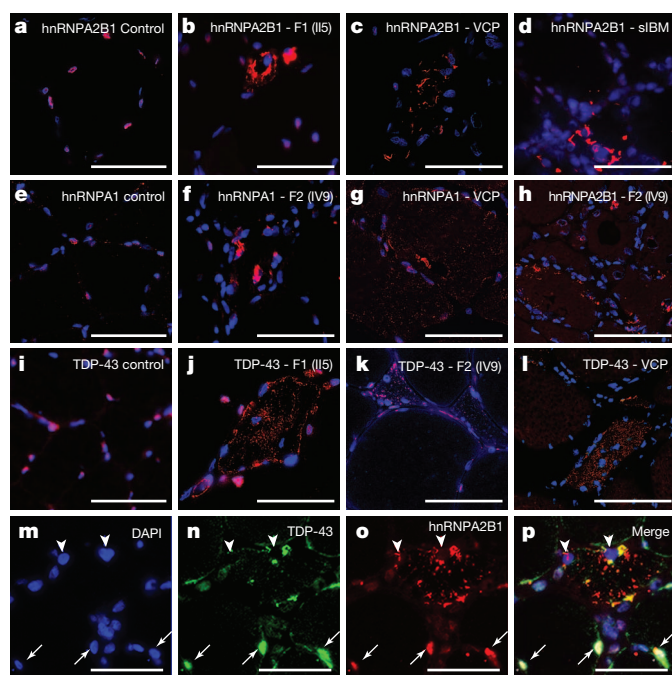


Figure 2 | Cytoplasmic pathology of hnRNP A2B1 and hnRNP A1. **a–d**, Immunohistochemical analysis of hnRNP A2B1 (red) in normal muscle (control, **a**), muscle biopsy from patient II5 from family 1 (**b**), a patient with MSP caused by a VCP mutation (R155H) (**c**) and a patient with sporadic IBM (sIBM, **d**). hnRNP A2B1 (red) was cleared from 4',6-diamidino-2-phenylindole (DAPI)-stained nuclei (blue) and accumulated in cytoplasmic inclusions (**b–d**). **e–g**, Immunohistochemical analysis of hnRNP A1 (red) in a normal muscle (**e**), a muscle biopsy from patient IV9 from family 2 (**f**) and a patient with MSP caused by a VCP mutation (R155H) (**g**). hnRNP A1 was cleared from nuclei and accumulated in cytoplasmic inclusions (**f, g**). **h**, In muscle tissue from patient IV9 (family 2), hnRNP A2B1 was cleared from nuclei and accumulated in cytoplasmic inclusions. (**i–l**) Immunohistochemical analysis of TDP-43 (red) in a normal muscle (**i**), muscle biopsy from patient II5 from family 1 (**j**), muscle biopsy from patient IV9 from family 2 (**k**) and a patient with MSP caused by a VCP mutation (R155H) (**l**). TDP-43 (red) was cleared from DAPI-stained nuclei (blue) and accumulated in cytoplasmic inclusions (**j–l**). **m–p**, Immunohistochemical analysis of hnRNP A2B1 and TDP-43 co-localization in a patient with sporadic IBM. DAPI (**m**), TDP-43 (**n**), hnRNP A2B1 (**o**) and merged images (**p**) are shown. hnRNP A2B1 (red) and TDP-43 (green) were cleared from nuclei (arrowheads) in an atrophic muscle fibre, but found in nuclei of neighbouring unaffected fibres (arrows). Scale bars, 50 μ m.

does not affect all hnRNPs. Double staining revealed that fibres with TDP-43 pathology typically also showed hnRNP A2B1 and hnRNP A1 pathology (Fig. 2m–p and Supplementary Fig. 4g, k). In these instances there was often partial co-localization of TDP-43, hnRNP A2B1 and hnRNP A1. We also observed limited immunopositivity for ubiquitin and p62 (Supplementary Fig. 4l–o).

Disease mutations affect hnRNP A2B1 and A1 PrLDs

hnRNP A2B1 and hnRNP A1 each have carboxy-terminal, glycine-rich domains that are essential for activity and mediate interaction with TDP-43 (ref. 8). FoldIndex predicts these domains to be intrinsically unfolded (Fig. 3a, b). Interestingly, these low-complexity domains have an amino-acid composition similar to that of yeast prion domains. Indeed, a previously devised algorithm¹⁴ predicted that the C-terminal regions of hnRNP A2 (residues 185–341) and hnRNP A1 (residues 186–320) are prion like (Fig. 3a–d), whereas these domains narrowly missed the cutoff for prion propensity according to a different algorithm¹⁵. Interestingly, similar PrLDs are present in many hnRNPs, including TDP-43 and FUS (Supplementary Fig. 5 and Supplementary Table 3). The disease-causing mutations fall at the centre of the predicted PrLD,

and, according to both algorithms, are expected to enhance prion-like behaviour (Fig. 3a, b).

We also examined hnRNP A2B1 and hnRNP A1 with ZipperDB, a structure-based threading algorithm, which scores six-amino-acid segments or their propensity to form two self-complementary beta strands termed ‘steric zippers’ that form the spine of amyloid fibrils¹⁶. Hexapeptides with Rosetta energy lower than -23 kcal mol⁻¹ are predicted to form steric zippers, with lower energy predicting higher amyloidogenicity. ZipperDB predicted that disease mutations increase the potency of steric zipper motifs present in the PrLDs of hnRNP A2B1 and hnRNP A1 (Fig. 3c–f). This analysis is discussed in greater detail in Supplementary Fig. 6. Taken together, multiple algorithms predict that the PrLDs of hnRNP A2B1 and hnRNP A1 are intrinsically disordered, but poised to access higher order self-templating structures. Notably, the disease mutations are centred within these PrLDs and are predicted to promote transition to an ordered structure.

Disease mutations accelerate hnRNP fibrillization

We directly tested the predictions that (1) hnRNP A2 and hnRNP A1 are prone to fibrillization and (2) that this property is enhanced by disease-causing mutations. First, we experimentally assessed the ZipperDB prediction for the impact of disease mutations on steric zipper motifs found in hnRNP A2B1 and hnRNP A1. Notably, the synthetic mutant hexapeptides of hnRNP A2B1(D290V) (Asn-Tyr-Asn-Val-Phe-Gly) and hnRNP A1(D262V) (Ser-Tyr-Asn-Val-Phe-Gly) rapidly assembled into amyloid fibrils, as shown by thioflavin T fluorescence and electron microscopy, whereas the corresponding wild-type peptides did not, even after several weeks (Fig. 4a–d). Thus, the disease-causing mutations in hnRNP A2B1 and hnRNP A1 generate highly amyloidogenic hexapeptides precisely as predicted. The more potent steric zippers that result from disease mutations are likely to be important for two reasons. First, introduction of similarly potent steric zippers is sufficient to force fibril formation, even in model proteins that do not ordinarily fibrillize¹⁷. Second, in hnRNP A1 and hnRNP A2 these potent steric zippers are centred in the intrinsically disordered PrLD and are thus available to make intermolecular contacts and drive fibril formation.

We next assessed the fibrillization propensity of full-length proteins. We purified bacterially expressed, glutathione S-transferase-tagged wild-type and mutant hnRNP A2 and hnRNP A1 as soluble proteins under native conditions and assessed fibrillization by sedimentation analysis and electron microscopy. Wild-type hnRNP A2 and hnRNP A1 were both intrinsically aggregation prone with a lag phase of ~ 4 h (Fig. 4e, g). Electron microscopy revealed that wild-type hnRNP A2 and hnRNP A1 form fibrils (Fig. 4f, h). The pronounced lag suggests that a critical rate-limiting step in hnRNP A2 and hnRNP A1 fibrillization is nucleation. Indeed, a small quantity (5% wt/wt) of preformed hnRNP A2 or hnRNP A1 fibrils greatly accelerated the assembly of hnRNP A2 or hnRNP A1, respectively (Supplementary Fig. 7a, e). By contrast, preformed hnRNP A2 fibrils did not seed assembly of hnRNP A1 (Supplementary Fig. 7e). Likewise, pre-formed hnRNP A1 fibrils did not seed assembly of hnRNP A2 (Supplementary Fig. 7a). Thus, wild-type hnRNP A1 and hnRNP A2 spontaneously form self-seeding fibrils.

Importantly, the disease mutations greatly accelerated hnRNP A2 and hnRNP A1 fibrillization (Fig. 4e–h). In all cases, the lag phase was curtailed and fibrillization was well advanced while the wild-type protein remained in lag phase. Thus, the disease-causing mutations directly promote nucleation of hnRNP A2 and hnRNP A1 into fibrils. Moreover, fibrils formed by hnRNP A2(D290V), hnRNP A1(D262V) and hnRNP A1(D262N) not only seeded their own assembly (Supplementary Fig. 7b, f, g), but also promoted fibrillization of their respective wild-type counterparts (Supplementary Fig. 7d, i). Neither hnRNP A2(D290V) fibrils nor hnRNP A1(D262V) fibrils seeded the assembly of TDP-43, another RNA-binding protein with a PrLD.

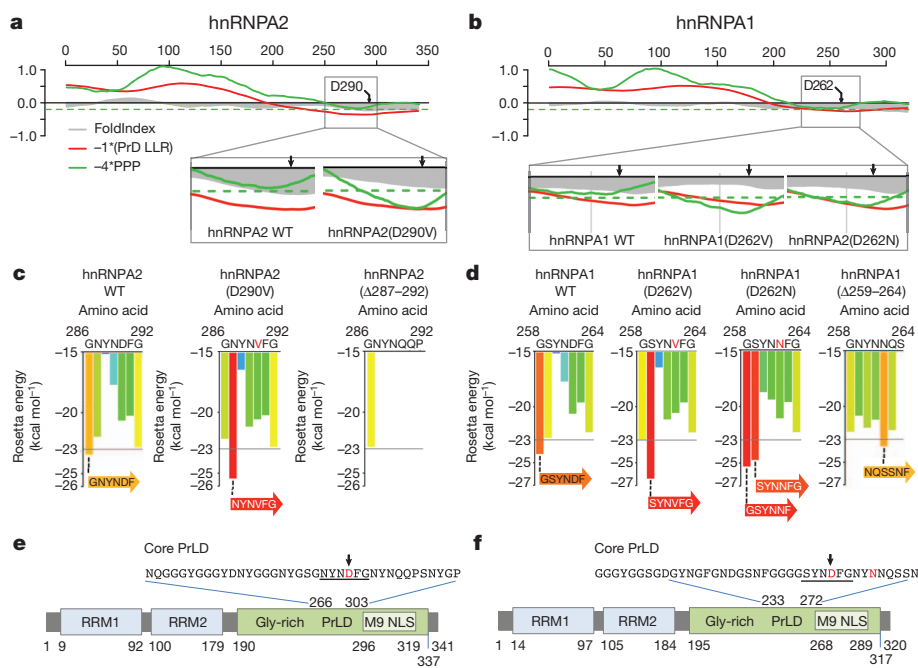


Figure 3 | The disease mutations affect a PrLD in hnRNP2B1 and hnRNP1. **a, b,** FoldIndex predicts an extended intrinsically unfolded region (grey curve less than zero) in the C-terminal region of hnRNP2 (**a**) and hnRNP1 (**b**). These regions were also predicted to be prion-like according to their log-likelihood ratio scores (PrD LLR) from the algorithm described in ref. 14 (red curve less than zero), and narrowly missed the cutoff for the predicted prion propensity (PPP) log-odds ratio scores from the algorithm described in ref. 15 (green curve below the dashed green line). All curves represent averages of 41 consecutive windows of 41 amino acids, corresponding to the criteria described in ref. 15. The disease mutations were predicted to make these domains more prionogenic (insets). **c, d,** ZipperDB-detected six-amino-acid stretches (underlined in **e, f**) within the core PrLDs for which the disease mutations increased the predicted amyloid-fibril-forming potential beyond the Rosetta threshold. **e, f,** Domain architecture of hnRNP2 (**e**) and hnRNP1 (**f**) shows the RNA-recognition motifs 1 and 2 (RRM1 and RRM2), the C-terminal glycine-rich domain and an M9 nuclear-localization signal. The PrLDs are centred in the C-terminal glycine-rich domain. Highly similar predictions were made for the minor isoforms of hnRNP2B1 (hnRNP1) and hnRNP1 (hnRNP1 isoform b). WT, wild type.

Finally, we assessed the importance of the steric zipper motifs in fibrillization. We deleted the steric zipper residues 287–292 from hnRNP2 and 259–264 from hnRNP1 and assessed fibrillization *in vitro*. Importantly, neither hnRNP2(Δ 287–292) nor hnRNP1(Δ 259–264) formed fibrils (Fig. 4e, g), even after 24 h of agitation. Moreover, hnRNP2(Δ 287–292) did not fibrillize when seeded by fibrils of wild-type or mutant hnRNP2 (Supplementary Fig. 7c), nor did hnRNP1(Δ 259–264) fibrillize when seeded by fibrils of wild-type or mutant hnRNP1 (Supplementary Fig. 7h). Thus, residues 287–292 of hnRNP2 and 259–264 of hnRNP1 are critical for spontaneous and seeded fibrillization of the full-length protein. Collectively, these findings indicate that steric zipper motifs centred in the PrLDs of wild-type hnRNP2 and hnRNP1 are critical to their intrinsic tendency to fibrillize and that disease mutations introduce more-potent steric zippers, which accelerate nucleation and polymerization (Supplementary Fig. 6).

Mutant hnRNP2B1 PrLD supports yeast prion activity

Yeast prion proteins are generally modular, meaning that prion domains from one protein can be transferred to another, while retaining prion activity¹⁸. To determine whether the PrLD of hnRNP2B1 could support prion activity in yeast, we replaced the Sup35 nucleation domain (residues 3–40) with the core PrLD (residues 261–303) from either wild-type or mutant hnRNP2B1 and expressed these fusion proteins as the sole copies of Sup35 in the cell. Prion formation was detected by monitoring nonsense suppression of the *ade2-1* allele and verified by monitoring a series of prion-defining criteria, including curability with guanidine, transferability with cytoduction and sensitivity to expression levels (detailed in Supplementary Fig. 8). This analysis showed that the hnRNP2B1 core PrLD can substitute for the Sup35 nucleation domain in supporting prion formation, and that the D290V mutation specifically promotes the nucleation activity. Moreover, full-length hnRNP1 and hnRNP2 form cytoplasmic aggregates and are toxic in yeast (Supplementary Fig. 9).

Mutations enhance hnRNP recruitment to RNA granules

Remarkably, the PrLDs in hnRNP2B1 and hnRNP1 correspond to the ‘low-complexity sequences’ found in various hnRNPs (including TDP-43 and FUS) that are essential determinants of RNA granule

assembly¹⁹. Stress granules are cytoplasmic ribonucleoprotein granules composed of repressed translation complexes^{20,21}. TDP-43 and FUS are recruited to stress granules and this is enhanced by disease mutations²⁰. Thus, we proposed that hnRNP2B1 and hnRNP1 would be recruited to stress granules and that this would be enhanced by disease mutations. To test this, we first examined the subcellular localization of endogenous hnRNP2B1 in HeLa cells before and after stress granule induction with arsenite. At baseline, endogenous hnRNP2B1 was localized exclusively to nuclei (Supplementary Fig. 10). After arsenite treatment hnRNP2B1 re-localized to eIF4G-positive cytoplasmic puncta, showing the recruitment of endogenous hnRNP2B1 to stress granules (Supplementary Fig. 10).

To examine the impact of disease mutations, we expressed Flag-tagged versions of wild-type and mutant hnRNP2. There was significantly greater incorporation of mutant hnRNP2 into constitutive stress granules than wild-type hnRNP2 (Fig. 5a). Moreover, after arsenite treatment mutant hnRNP2 was incorporated into stress granules more rapidly than wild-type hnRNP2, although by 30 min their degree of incorporation was similar (Fig. 5b). Both wild-type and mutant hnRNP2 formed high-molecular-weight species that accumulated in the detergent-insoluble fraction (Fig. 5c). Similar to the kinetics of stress granule incorporation, mutant hnRNP2 accumulated as high-molecular-weight species more rapidly than wild-type hnRNP2, although by 30 min their accumulation was comparable (Fig. 5c). Similar results were obtained upon examination of endogenous hnRNP1 as well as exogenous wild-type and mutant hnRNP1 (Supplementary Fig. 11). Interestingly, hnRNP2-positive and hnRNP1-positive stress granules were also immunopositive for TDP-43 (Supplementary Fig. 12). We also generated fibroblast cell lines from patients II4 and II6 (family 1). In these cells endogenous mutant hnRNP2 accumulated in constitutive stress granules that were immunopositive for eIF4G in addition to TDP-43 and VCP (Supplementary Fig. 13).

Mutant hnRNPs form pathological inclusions *in vivo*

We generated transgenic *Drosophila* expressing wild-type or mutant forms of human hnRNP2 and hnRNP1, as well as wild-type or mutant forms of the fly homologue HRB98DE (Supplementary Figs 14 and 15). Expression of wild-type hnRNP2 in *Drosophila*

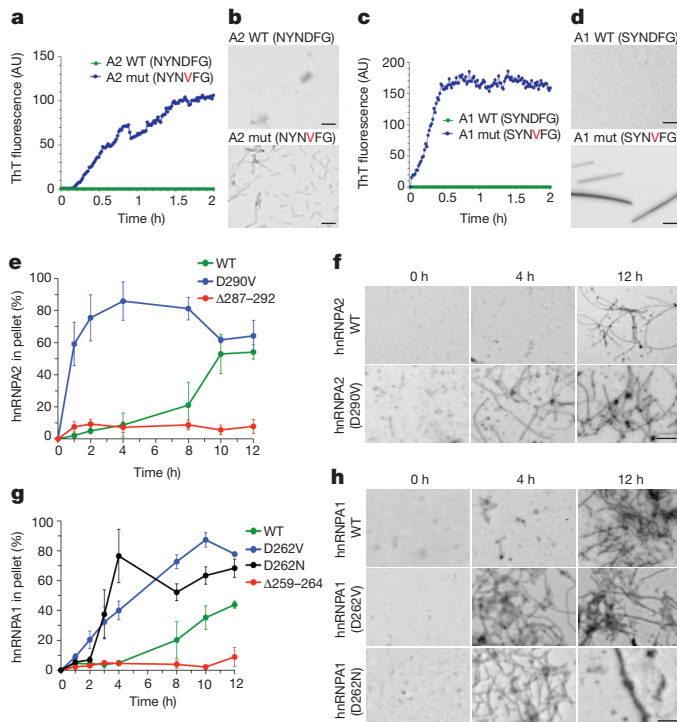


Figure 4 | Disease mutations accelerate hnRNP A2 and hnRNP A1 fibrillization. **a**, Synthetic hexapeptides A2 wild-type (Asn-Tyr-Asn-Asp-Phe-Gly) or mutant (Asn-Tyr-Asn-Val-Phe-Gly) were incubated at 25 °C for 2 h. Fibrillization was monitored by thioflavin T (ThT) fluorescence. **b**, Electron micrograph of A2 wild-type or mutant hexapeptides after 10 min at 25 °C. Scale bars, 0.1 μm. **c**, **d**, Fibrillization analysis of A1 wild-type (Ser-Tyr-Asn-Asp-Phe-Gly) or mutant (Ser-Tyr-Asn-Val-Phe-Gly) as in **a**, **b**. Scale bars, 0.5 μm. **e**, Full-length hnRNP A2 wild-type, hnRNP A2(D290V) or hnRNP A2(Δ287–292) was incubated at 25 °C with agitation for 0–12 h. At various times, the amount of aggregated hnRNP A2 was determined. Values represent means ± s.e.m. ($n = 3$). **f**, Electron micrograph of hnRNP A2 fibrillization reactions after 0, 4 and 12 h at 25 °C. Note the absence of fibres after 4 h for wild-type hnRNP A2. Scale bar, 0.5 μm. **g**, **h**, Fibrillization of full-length hnRNP A1 wild-type, hnRNP A1(D262V), hnRNP A1(D262N) or hnRNP A1(Δ259–264), monitored as in **e**, **f**. Scale bar, 0.5 μm.

indirect flight muscle led to mild degeneration affecting the rostral portion of several muscles, whereas expression of hnRNP A2(D290V) caused severe degeneration that affected all muscles (Fig. 6a). Importantly, flies overexpressing hnRNP A2(Δ287–292), in which the hexapeptide (Asn-Tyr-Asn-Asp/Val-Phe-Gly) in the PrLD was deleted, had virtually normal muscles (Fig. 6a). This result indicates that the severe toxicity found in flies overexpressing mutant hnRNP A2 requires the presence of this potent steric zipper motif.

Immunohistochemical analysis showed that wild-type hnRNP A2 localized appropriately to nuclei, as did hnRNP A2(Δ287–292), whereas hnRNP A2(D290V) largely accumulated in cytoplasmic inclusions (Fig. 6b). We observed a correlation between cytoplasmic inclusion formation and solubility of hnRNP A2. Specifically, mutant hnRNP A2(D290V) protein was largely recovered from the detergent-insoluble fraction, whereas wild-type hnRNP A2 and hnRNP A2(Δ287–292) proteins were mostly found in the detergent-soluble fraction (Fig. 6c, d). Thus, the degree of muscle degeneration in flies expressing hnRNP A2 correlates with the extent of cytoplasmic inclusions and with hnRNP A2 solubility. Entirely consistent results were obtained upon expression of wild-type and mutant versions of hnRNP A1 and HRB98DE (Supplementary Fig. 15).

Last, we evaluated the impact of the disease mutation on hnRNP A2 cellular localization in mammalian muscle. We electroporated mouse tibialis anterior with plasmid expressing Flag-tagged versions of wild-type or mutant hnRNP A2. Wild-type hnRNP A2 was localized

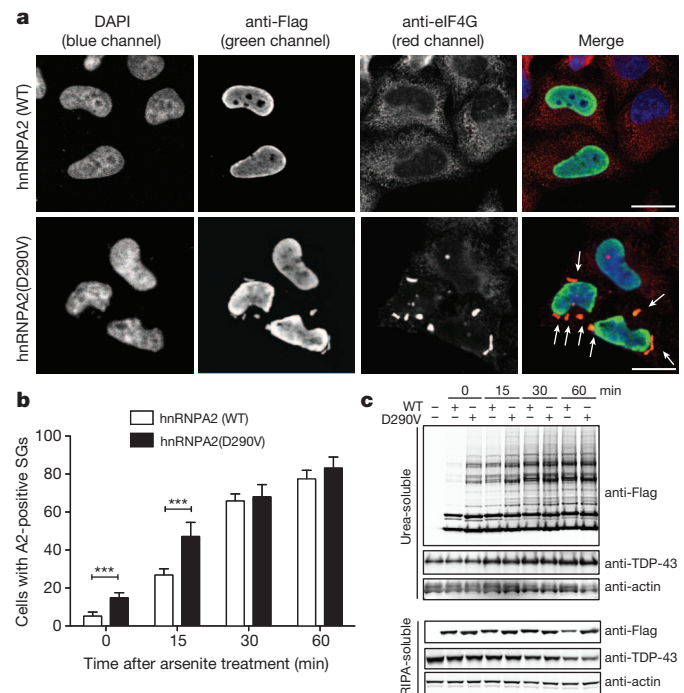


Figure 5 | hnRNP A2 recruitment to stress granules is accelerated by disease mutation. **a**, HeLa cells were transfected with Flag-tagged wild-type or mutant hnRNP A2 and stained with anti-Flag (green), anti-eIF4G (red) and DAPI (blue). Arrows indicate hnRNP A2- and eIF4G-positive stress granules. Scale bar, 20 μm. **b**, HeLa cells were transfected as in **a**, treated with 0.5 mM sodium arsenite for the indicated time, and immunostained as in **a**. The percentage of cells displaying hnRNP A2-positive stress granules (SGs) at indicated time points after treatment with arsenite cells are plotted. Data represent mean ± s.e.m. $n = 3$; *** $P < 0.001$. **c**, HeLa cells were transfected and stimulated as in **b**, and sequentially extracted with RIPA (radio immunoprecipitation assay) buffer and urea buffer. Immunoblotting was conducted with anti-Flag, anti-TDP-43 or anti-actin (loading control) antibodies.

appropriately within nuclei, whereas mutant hnRNP A2(D290V) was excluded from nuclei and accumulated in cytoplasmic inclusions (Supplementary Fig. 16), highly reminiscent of the pathology seen in MSP patients.

Discussion

Polymerization of PrLDs underlies the orderly, regulated phase transition that drives assembly of non-membrane-bound organelles including RNA granules, which serve as ‘reaction crucibles’ in which many aspects of RNA metabolism occur²². Disease mutations introduce more potent steric zippers into the PrLDs of hnRNP A2 and hnRNP A1, dysregulating and accelerating nucleation and polymerization, altering the dynamics of RNA granule assembly, which likely has adverse consequences for RNA metabolism. In addition to hnRNP A2 and hnRNP A1, at least four other RNA-binding proteins that harbour PrLDs (TDP-43, FUS, EWSR1 and TAF15) accumulate in disease pathology²³. By analogy to hnRNP A2B1 and hnRNP A1, we speculate that these and perhaps other PrLD-containing proteins contribute to the initiation or propagation of disease by a similar mechanism. Indeed, ~250 human proteins are predicted to harbour PrLDs, many of which are RNA-binding proteins (Supplementary Fig. 5 and Supplementary Table 3), and should be considered candidates for contributing to degenerative disease.

Thus, disease could ensue from unregulated polymerization initiated spontaneously by PrLDs upon environmental stress, or by a mutation affecting either the PrLD or another factor (for example, VCP) that regulates the ribonucleoprotein granule assembly with altered metabolism of RNA as a consequence.

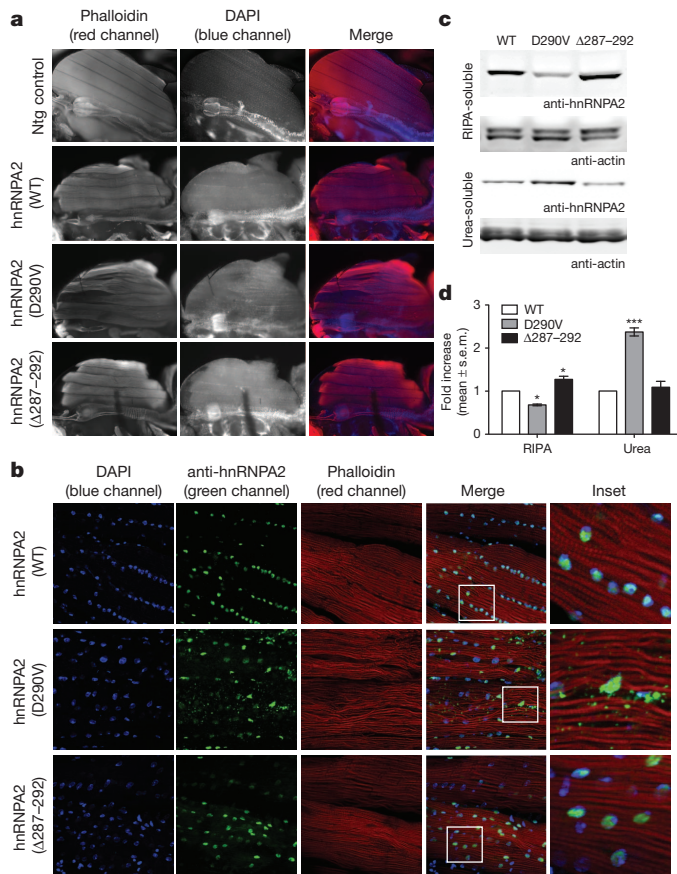


Figure 6 | Mutant hnRNP A2 forms cytoplasmic inclusions in *Drosophila*. **a**, Adult flies were dissected to expose the dorsal longitudinal indirect flight muscle and stained with Texas Red-phalloidin (red) and DAPI (blue). Flies expressing human wild-type hnRNP A2 under control of the *Mhc*-Gal4 driver showed mild degeneration, whereas flies expressing mutant human hnRNP A2 show severe degeneration affecting all muscles. Flies expressing hnRNP A2(Δ287–292) show muscle histology similar to flies expressing wild-type hnRNP A2. Ntg indicates non-transgenic control. **b**, Wild-type hnRNP A2 localizes exclusively to nuclei, whereas hnRNP A2(D290V) also accumulates extensively in cytoplasmic inclusions. hnRNP A2(Δ287–292) localizes exclusively to nuclei. **c**, Thoraces of adult flies were dissected and sequential extractions were performed to examine the solubility profile of hnRNP A2. **d**, Quantification of the blot shown in **c**. Data represent mean ± s.e.m. $n = 3$; * $P < 0.05$, *** $P < 0.001$; two-way analysis of variance with Bonferroni's post hoc test.

Notably, diseases associated with pathological inclusions of PrLD-containing proteins frequently show 'spreading' pathology, in which cellular degeneration with intracellular inclusions starts in one epicentre and subsequently spreads to neighbouring tissue²⁴. Although not directly addressed here, this study suggests that cell-to-cell transmission of a self-templating conformer could contribute to the spreading pathology that is characteristic of these diseases.

The discovery of VCP mutations in MSP¹ led to the subsequent discovery of pathogenic VCP mutations in sporadic or familial forms of ALS², FTD⁴, IBM⁵ and PDB⁶. Therefore, subsequent to the discovery of mutations in hnRNP A2B1 and hnRNP A1 underlying MSP, we screened 212 familial ALS cases and identified a pathogenic hnRNP A1 mutation in one family, as described earlier. As this manuscript was being assembled for publication we completed screening of 305 sporadic ALS cases and identified a non-synonymous variant in hnRNP A1 (c.800/956A>G, p.N267S/N319S) in one classic, late-onset case in which mutations in known ALS genes had been excluded (Supplementary Fig. 17). Although not definitive, this variant is probably pathogenic because it is centred in the core PrLD and introduces a potent steric zipper similar to that introduced by D262V/N mutations. The frequency of hnRNP A2B1 and hnRNP A1 mutations in

ALS, and the possibility that mutations in these genes underlie some sporadic and familial forms of FTD, IBM and PDB, will be important to address in the future.

METHODS SUMMARY

Exome enrichment of 3 µg genomic DNA was performed by using the Agilent SureSelect Human All Exon Kit capture library (G3362 for five DNA samples from family 1 and G3370 for five DNA samples from family 2) according to the manufacturer's protocol (SureSelect Human All Exon Illumina Paired-End protocol version 1.0.1). Captured DNAs were sequenced on an Illumina GAIIX sequencer (Illumina) with 76-base-pair (family 1) or 100-base pair (family 2) paired-end reads. At least two lanes of sequencing data were collected for each sample to generate sufficient coverage. Image analyses and base calling were performed by using the Illumina Genome Analyzer Pipeline software (GAPipeline version 1.5 or higher) with default parameters. Reads were aligned to a human reference sequence (University of California, Santa Cruz (UCSC) assembly hg19, National Center for Biotechnology Information (NCBI) build 37) and genotypes were called at all positions at which there were high-quality sequence bases (Phred-like score Q25 or greater) at minimum coverage of 8, using CLC Genomics Workbench v4.5.1 (CLC Bio). For each sample, more than 50% of at least 100 million reads were uniquely mapped to the targeted human exon regions to give mean depth of coverage of 123. For linkage analysis samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. Single nucleotide polymorphism genotype calls were performed according to the standard Affymetrix protocols. A subset of single nucleotide polymorphisms chosen on the basis of heterozygosity was used for linkage analysis, using dChip software (<http://www.hsph.harvard.edu/cli/complab/dchip/>) under parametric linkage analysis and a dominant model with 99% penetrance in heterozygotes and a disease allele frequency of 0.001.

Full Methods and any associated references are available in the online version of the paper.

Received 5 January 2012; accepted 17 January 2013.

Published online 3 March 2013.

- Nalbandian, A. *et al.* The multiple faces of valosin-containing protein-associated diseases: inclusion body myopathy with Paget's disease of bone, frontotemporal dementia, and amyotrophic lateral sclerosis. *J. Mol. Neurosci.* **45**, 522–531 (2011).
- Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68**, 857–864 (2010).
- Watts, G. D. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature Genet.* **36**, 377–381 (2004).
- Neumann, M., Tolnay, M. & Mackenzie, I. R. The molecular basis of frontotemporal dementia. *Expert Rev. Mol. Med.* **11**, e23 (2009).
- Shi, Z. *et al.* Characterization of the Asian myopathy patients with VCP mutations. *Eur. J. Neurol.* **19**, 501–509 (2012).
- Chung, P. Y. *et al.* Indications for a genetic association of a VCP polymorphism with the pathogenesis of sporadic Paget's disease of bone, but not for TNFSF11 (RANKL) and IL-6 polymorphisms. *Mol. Genet. Metab.* **103**, 287–292 (2011).
- Kottlors, M. *et al.* Late-onset autosomal dominant limb girdle muscular dystrophy and Paget's disease of bone unlinked to the VCP gene locus. *J. Neurol. Sci.* **291**, 79–85 (2010).
- Buratti, E. *et al.* TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J. Biol. Chem.* **280**, 37572–37584 (2005).
- Ritson, G. P. *et al.* TDP-43 mediates degeneration in a novel *Drosophila* model of disease caused by mutations in VCP/p97. *J. Neurosci.* **30**, 7729–7739 (2010).
- Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256–271 (2006).
- Sofola, O. A. *et al.* RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CCG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS. *Neuron* **55**, 565–571 (2007).
- Jin, P. *et al.* Pur α binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome. *Neuron* **55**, 556–564 (2007).
- Salajegheh, M. *et al.* Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis. *Muscle Nerve* **40**, 19–31 (2009).
- Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).
- Toombs, J. A., McCarty, B. R. & Ross, E. D. Compositional determinants of prion formation in yeast. *Mol. Cell. Biol.* **30**, 319–332 (2010).
- Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **107**, 3487–3492 (2010).
- Teng, P. K. & Eisenberg, D. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng. Des. Sel.* **22**, 531–536 (2009).

18. Li, L. & Lindquist, S. Creating a protein-based element of inheritance. *Science* **287**, 661–664 (2000).
19. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
20. Wolozin, B. Regulated protein aggregation: stress granules and neurodegeneration. *Mol. Neurodegener.* **7**, 56 (2012).
21. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36**, 932–941 (2009).
22. Weber, S. C. & Brangwynne, C. P. Getting RNA and protein in phase. *Cell* **149**, 1188–1191 (2012).
23. Neumann, M. *et al.* FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* **134**, 2595–2609 (2011).
24. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* **1462**, 61–80 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the patients whose participation made this work possible. We thank the St Jude Pediatric Cancer Genome Project and J. Zhang in particular for providing access to control sequencing data. We thank C. Gellera,

B. Baloh, M. Harms, S. Krause, G. Dreyfuss and T. Cundy for sharing reagents. We thank S. Donkervoort and S. Mumm for coordinating samples, and A. Taylor for editorial assistance. J.P.T. was supported by ALSAC, the Packard Foundation and the National Institutes of Health (NIH) (NS053825); J.P.T. and M.B. were supported by the ALS Association; J.Q.T. was supported by the NIH (AG032953); J.S. was supported by the NIH (DP20D002177 and NS067354) and the Ellison Medical Foundation; E.D.R. was supported by the National Science Foundation (MCB-1023771). C.C.W. was supported by the NIH (AG031867).

Author Contributions H.J.K., N.C.K., E.D.R., C.C.W., J.S. and J.P.T. designed experiments. H.J.K., N.C.K., E.A.S., J. Moore, Z.D., K.S.M., B.F., S.L., A.M., A.P.K., Y.R.L. and A.F.F. performed the experiments. K.B.B., A.M.W., R.R., J.L.P., S.A.G., J.Q.T., B.N.S., S.T., A.-S.G., J. Miller, C.E.S., M.K., J.K., A.P., M.B. and V.E.K. provided patient clinical material, clinical evaluation, or evaluation of patient clinical material. H.J.K., N.C.K., Y.-D.W., R.C., B.J.T., A.D.G., O.D.K., E.D.R., J.S. and J.P.T. contributed to data analysis. E.D.R., O.D.K. and C.C.W. contributed to manuscript preparation. H.J.K., J.S. and J.P.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P.T. (JP.Taylor@STJUDE.ORG) or J.S. (jshorter@mail.med.upenn.edu).

METHODS

Patients. Patients were examined by M.B. and J.P.T. except where indicated. The Institutional Review Board of the University of Miami approved the study protocol and all participants provided written informed consent.

Exome capture and next-generation sequencing. Exome enrichment of 3 µg genomic DNA was performed by using the Agilent SureSelect Human All Exon Kit capture library (G3362 for five DNA samples from family 1 and G3370 for five DNA samples from family 2) according to the manufacturer's protocol (SureSelect Human All Exon Illumina Paired-End protocol version 1.0.1). The capture library, containing regions totalling approximately 38 megabases (G3362) or 50 megabases (G3370), is designed to target all human exons in the NCBI consensus coding sequence (CCDS) database. Captured DNAs were sequenced on an Illumina GAIIX sequencer (Illumina) with 76-base-pair (family 1) or 100-base-pair (family 2) paired-end reads. At least two lanes of sequencing data were collected for each sample to generate sufficient coverage. Image analyses and base calling were performed by using the Illumina Genome Analyzer Pipeline software (GAPipeline version 1.5 or higher) with default parameters. Reads were aligned to a human reference sequence (UCSC assembly hg19, NCBI build 37) and genotypes were called at all positions at which there were high-quality sequence bases (Phred-like score Q25 or greater) at minimum coverage of 8, using CLC Genomics Workbench v4.5.1 (CLC Bio). For each sample, more than 50% of at least 100 million reads were uniquely mapped to the targeted human exon regions to give a mean depth of coverage of 123. Under such coverage, approximately 94% of targeted regions were covered by 5 reads or more, and more than 85% were covered by more than 20 reads (Supplementary Fig. 1). To identify the pathogenic mutations, we performed a series of filtering steps (Supplementary Fig. 1). We first discarded the variants that did not change the amino-acid sequence. We then generated a list of variants that are present in affected individuals but absent in unaffected individuals. Variants previously reported in the Single Nucleotide Polymorphism Database build 132 (dbSNP132) hosted by the National Center for Biotechnology Information were excluded. The variants were further filtered by using the in-house database (a collection of variants from whole-exome sequencing of more than 625 individuals) by a similar methodology. Finally Sanger sequencing was performed on DNA of three additional members (both affected and unaffected) from each family to exclude the nonpathogenic mutations. Presence of mutations was confirmed in all candidate disease-causing variants by Sanger sequencing.

Linkage analysis. Eight DNA samples (parents and six siblings) from family 1 were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. SNP genotype calls were performed according to the standard Affymetrix protocols. A subset of SNPs chosen on the basis of heterozygosity was used for linkage analysis, using dChip software (<http://www.hsph.harvard.edu/cli/complab/dchip>) under parametric linkage analysis and a dominant model with 99% penetrance in heterozygotes and a disease allele frequency of 0.001. Five genomic segments with a size greater than 100 kilobases, at least 20 SNP markers, and a log odds ratio (lod) score greater than 0 are shown in Supplementary Fig. 1.

Generation of phylogram for hnRNP A/B gene family. TBPH of *D. melanogaster* was used to BLAST (basic local alignment search tool) against the RefSeq protein database including species of *M. musculus*, *D. melanogaster*, *A. gambiae*, *H. sapiens*, *D. rerio* (zebrafish), *C. elegans* and *G. gallus* (jungle fowl). The top 100 sequences (according to BLAST *E* values) were obtained. Sequences sharing more than 98% similarity were filtered by UCLUST²⁵ because they likely represent alternatively spliced variants of the same genes. Eighty-one sequences remaining after this process were aligned using the MAFFT multiple sequence alignment program²⁶. The DNA-coding sequences of all proteins were extracted and threaded onto the protein sequences to form an in-frame codon alignment. Squint²⁷ was used to extract the DNA-coding alignment of the two RRM domains present in each protein using the protein alignment as a guide. Phylogenetic trees of the codon alignment including 21 genes from only human and *Drosophila* were generated using the genetic algorithm-based GARLI²⁸ method.

Constructs. Complementary DNAs encoding human hnRNP2 (accession number NM_002137.2) or hnRNP1 (accession number NM_002136.1) in the plasmid pCMV6-XL5 were obtained from OriGene. A Flag epitope tag was added to the 5' end of hnRNP2 or hnRNP1 by PCR. The cDNAs were then subcloned into a pcDNA 3.1 plasmid, using restriction sites EcoRI and XhoI, creating pcDNA 3.1-Flag-hnRNP2 or -Flag-hnRNP1. cDNAs encoding *Drosophila* HRB98DE (accession number NM_170373.1) were obtained from the *Drosophila* Genomics Research Center. Missense mutations of hnRNP2 (hnRNP2 (D290V)), hnRNP1 (hnRNP1 (D262V)) and *Drosophila* HRB98DE (HRB98DE (D302V)) were generated by site-directed mutagenesis. Sequence analysis was conducted on all plasmid constructs.

Antibodies. The following commercial antibodies were used in this study: mouse monoclonal anti-hnRNP2B1 antibodies EF-67 and DP3B3 (Santa

Cruz Biotechnology), rabbit polyclonal anti-TDP-43 antibody (Protein Tech Group), anti-eIF4G antibody (Santa Cruz Biotechnology), anti-hnRNP1 antibody (Cell Signaling Technology), anti-VCP antibody (Epitomics) and anti-Flag M2 antibody (Sigma). Anti-hnRNP1 antibodies (4B10 and 9H10) were generous gifts from G. Dreyfuss. Validation of these antibodies is shown in Supplementary Fig. 18.

Cell culture and transfection. HEK293 cells and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 1% L-glutamate. Human fibroblasts were grown in DMEM nutrient mixture F-12 (DMEM/F12) supplemented with 20% FBS, 1% penicillin/streptomycin and 1% L-glutamate. Cells were transfected using Lipofectamine LTX with Plus Reagent (Invitrogen), according to the manufacturer's instructions.

RNA interference. Short interfering RNAs (siRNAs) against hnRNP2B1 were ordered from Santa Cruz Biotechnology (sc-43841, target sequences: 5'-CAGUCCGUAAGCUCUUUAtt, GGAUGGCUAUAUUGGGUAUtt-3' and 5'-GGAUCAUGGUGUAAUAAGAtt-3') and Dharmacon (ON-TARGETplus SMARTpool, human HNRPA2B1 (3181), target sequences: 5'-CGGUGGA AAUUCGGACCA-3', 5'-GCUGUUUGUUGCGGAAU-3', 5'-GGAGAG UAGUUGAGCCAAA-3' and 5'-GAGGAGGAUCUGAUGGAUA-3'). Lower-case letters indicate non-targeting 3'-overhang included for stability. Control siRNAs were ordered from Dharmacon (ON-TARGETplus Non-targeting siRNA no. 1). HeLa cells were transfected in six-well plates with Lipofectamine RNAiMax (Invitrogen) according to the manufacturer's protocol and collected after 48 and 72 h.

Immunofluorescence studies. Cells were fixed in 4% paraformaldehyde in PBS buffer, permeabilized with 0.5% Triton X-100 in PBS for 10 min, blocked with 5% goat serum in PBS for 30 min and incubated with primary antibody for 2 h at room temperature (20–25 °C) or overnight at 4 °C. Primary antibodies were visualized with secondary antibodies conjugated with Alexa Fluor 488 and Alexa Fluor 594 (Molecular Probes, Invitrogen) and nuclei were detected using DAPI. Stained cells were examined using a confocal microscope (Zeiss LSM 510 NLO Meta) with Zeiss ZEN software.

Solubility and biochemical analyses. Sequential extractions were performed to examine the solubility profile of hnRNP2B1. Cells were washed twice with PBS, lysed in cold RIPA buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS and 1 mM EDTA) and sonicated. Cell lysates were cleared by centrifugation at 100,000g for 30 min at 4 °C to generate RIPA-soluble samples. To prevent carry-overs, the resulting pellets were washed twice with PBS (that is, re-sonicated and re-centrifuged). Only supernatants from the first centrifugation were analysed. RIPA-insoluble pellets were then extracted with urea buffer (7 M urea, 2 M thiourea, 4% CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulphonate), 30 mM Tris, pH 8.5), sonicated and centrifuged at 100,000g for 30 min at 22 °C. Protease inhibitors were added to all buffers before use. Protein concentration was determined by the bicinchoninic acid method (Pierce) and proteins were resolved by NuPAGE Novex 4–12% Bis-Tris Gel (Invitrogen).

In vivo electroporation of mice. All animal work was performed with the approval of the St Jude Children's Research Hospital Committee on Animal Care. Mice were anesthetized and the skin overlying the tibialis anterior muscle was shaved, and the animals were injected with endotoxin-free Flag-tagged wild-type hnRNP2 or Flag-tagged hnRNP2(D290V) mutant diluted in sterile PBS to a volume of 50 µl by using a 0.5-ml syringe fitted with a 29-gauge needle. Two-needle array electrodes (BTX Technologies) were inserted into the muscle immediately after DNA delivery for electroporation. The distance between the electrodes was 5 mm, and the array was inserted longitudinally relative to the muscle fibres. *In vivo* electroporation parameters were as follows: voltage, 75 V; pulse length, 50 ms; number of pulses, 6; pulse interval, 200 ms; desired field strength, 200 V cm⁻¹, given by a BTX ECM830 Electro Square Porator. After 7 days of recovery, mice were processed for immunofluorescence analysis.

Muscle histochemistry and immunohistochemistry. Human samples were processed for routine histochemical analysis as previously described¹³. For skeletal muscle immunohistochemistry, isolated skeletal muscle was mounted by using tragalanth gum and quick-frozen in liquid-nitrogen-cooled 2-methylbutane. Sections (8 µm) were placed on slides and fixed in 4% paraformaldehyde for 5 min and then ice-cold acetone for 5 min. Sections were incubated with PNB blocking reagent (Perkin Elmer) for 1 h and then incubated overnight in primary antibody diluted in PNB. After serial washes in 1× PBS, slides were incubated in secondary antibody diluted in PNB for 2 h. Sections were mounted with Mowiol containing DAPI. Specimens were examined using a fluorescent microscope (80i upright; Nikon) and charge-coupled device camera (EZ monochrome; Roper Industries) with deconvolution software analysis (NIS Elements; Nikon). Nonfluorescent images were taken with a 5-megapixel colour charge-coupled

device (Nikon). Image processing and analysis were performed with NIS Elements 4.0 software and Photoshop CS3 (Adobe).

Yeast strains, media and plasmids. Yeast cells were grown in rich media (YPD) or in synthetic media lacking uracil and containing 2% glucose (SD/–Ura), raffinose (SRaf/–Ura) or galactose (SGal/–Ura). The TDP-43 and FUS yeast expression constructs have been previously described^{29–31}. Wild-type and IBMPFD-associated mutant hnRNA2B1 and hnRNA1 Gateway entry clones were generated by PCR, incorporating the flanking Gateway attB1 and attB2 sites along with a Kozak consensus sequence. Resulting PCR products were shuttled into pDONR221, using a Gateway BP Clonase reaction. Two versions of each entry clone (with or without stop codon) were then used in LR reactions with pAG416Gal-cdbB, pAG426Gal-cdbB, pAG416Gal-cdbB-GFP or pAG426Gal-cdbB-GFP³² to generate the CEN (centromeric) and 2- μ m untagged or YFP-fusion constructs. Primer sequences are available upon request. To prevent unwanted recombination events owing to unstable repetitive sequences in the hnRNA2B1 and hnRNA1 DNA sequence, we propagated plasmids in SURE-2 or STBL-3 *Escherichia coli* at 30 °C. For testing the phenotype of wild-type and IBMPFD-associated mutant hnRNA2 as Sup35-fusion proteins, PCR was used to amplify amino acids 261–303 of hnRNA2 and insert it into Sup35 in the place of amino acids 3–40. The resulting PCR products were co-transformed with BamHI/HindIII-cut pJ526 (from D. Masison, NIH; ref. 33) into strain 780-1D (*MAT α kar1-1 SUQ5 ade2-1 his3 Δ 202 leu2 Δ 1 trp1 Δ 63 ura3-52 sup35::KanMX*) containing the SUP35 maintainer plasmid pJ533 (from D. Masison; ref. 34). Transformants were selected on medium lacking leucine, and then transferred to 5-fluoroorotic-acid-containing medium to select for loss of pJ533. Sequences were confirmed by DNA sequencing. Overexpression plasmids were generated by PCR amplification, restriction digest and cloning into compatible sites in pKT24 (from K. Taylor, NABI), which contains a *GAL1* promoter, as previously described³⁵.

Yeast transformation and spotting assays. Yeast procedures were performed according to standard protocols³⁶. The polyethylene glycol/lithium acetate method was used to transform yeast with plasmid DNA³⁷. For spotting assays, yeast cells were grown overnight at 30 °C in SRaf/–Ura until they reached log or mid-log phase. Cultures were then normalized for $D_{600\text{ nm}}$, serially diluted, spotted onto synthetic solid media containing SD/–Ura or SGal/–Ura, and grown at 30 °C for 2–3 days. hnRNA2B1–Sup35 fusions were tested for prion formation and for stability and curability of the Ade⁺ phenotype as previously described³⁵. Cytoductions were performed as previously described³⁸. The recipient strain was YER746 (*MAT α kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 arg1::HIS3 sup35::KanMX*), carrying plasmid pER697 (URA3), which expresses the hnRNA2(D290V)–Sup35 fusion from the SUP35 promoter.

Fluorescence microscopy for yeast. For fluorescence microscopy experiments, single-colony isolates of yeast strains were grown to mid-log phase in SRaf/–Ura media at 30 °C. Cultures were spun down and re-suspended in the same volume of SGal/–Ura to induce expression of the hnRNA2B1 or hnRNA1 constructs. Cultures were induced with galactose for 4–6 h and processed for microscopy. Images were obtained by an Olympus IX70 inverted microscope and a Photometrics CoolSnap HQ 12-bit CCD camera. Z stacks of several fields were collected for each strain. The images were de-blurred using a nearest neighbour algorithm in the DeltaVision softWoRx software and representative cells were chosen.

Hexapeptide assembly. The wild-type hnRNA2 (Asn-Tyr-Asn-Asp-Phe-Gly), mutant hnRNA2(D290V) (Asn-Tyr-Asn-Val-Phe-Gly), wild-type hnRNA1 (Ser-Tyr-Asn-Asp-Phe-Gly) and mutant hnRNA1(D262V) (Ser-Tyr-Asn-Val-Phe-Gly) hexapeptides were synthesized at the Keck Biotechnology Resource Laboratory at Yale University School of Medicine. A1 peptides were dissolved at 1 mM and A2 peptides at 5 mM in 150 mM KCl, 40 mM HEPES-KOH, pH 7.4 and 1 mM dithiothreitol and used immediately for assembly reactions. Fibre-assembly reactions were monitored at $\lambda_{\text{ex/em}}$ 440/482 nm, in the presence of 25 μ M thioflavin T, at room temperature on a Tecan Safire² or Tecan Infinite M1000 plate reader. Alternatively, reactions were processed for electron microscopy as for the full-length hnRNPs (see below).

hnRNP purification. Wild-type and mutant hnRNA2 or hnRNA1 were expressed and purified from *E. coli* as GST-tagged proteins. Expression constructs were generated in pDuet to contain a TEV-cleavable site, resulting in a GST-TEV-hnRNP construct^{30,31}. GST-TEV-hnRNP was overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells (Agilent) and purified under native conditions using a glutathione-sepharose column (GE) according to the manufacturer's instructions. Proteins were eluted from the glutathione sepharose with assembly buffer (AB; 40 mM HEPES-KOH, 150 mM KCl, 5% glycerol, 20 mM glutathione, pH 7.4). Protein was centrifuged for 10 min at 16,100g, and supernatant was separated from pellet to remove any protein aggregates. Protein concentration was determined by Bradford assay (Bio-Rad) in comparison to BSA standards.

In vitro fibril-formation assay. For GST-tagged protein, aggregation was initiated by adding TEV protease (Invitrogen) to GST-TEV-hnRNA2 (3 μ M) or GST-TEV-hnRNA1 (5 μ M) in AB. Aggregation reactions were incubated at 25 °C for 0–12 h with or without agitation at 1,200 r.p.m. in an Eppendorf Thermomixer. For self-seeded and cross-seeded reactions, hnRNA2, hnRNA2(D290V), hnRNA1 or hnRNA1(D262V) fibrils were assembled at 5 μ M for 24 h with agitation at 1,400 r.p.m. in AB. These preformed fibrils then used to seed (5% wt/wt) the assembly of hnRNA2 (2.5 μ M), hnRNA2(D290V) (2.5 μ M), hnRNA1 (5 μ M), hnRNA1(D262V) (5 μ M) or hnRNA1(D262N) (5 μ M) in AB as indicated. Here, assembly reactions were agitated at 1,200 r.p.m. at 25 °C. For hnRNA2 cross-seeding reactions the amount of preformed fibrils added was increased to 10% wt/wt and reactions were not agitated. In hnRNA1 cross-seeding reactions, preformed fibrils were briefly sonicated to fragment fibrils before addition to the assembly reaction. For hnRNA1 cross-seeding reactions the amount of preformed fibrils added was 5% wt/wt and reactions were not agitated. For sedimentation analysis, samples were centrifuged at 16,100g for 10 min at 25 °C. Pellet fractions were resolved by SDS-PAGE and stained with Coomassie blue. The amount of protein in the pellet fraction was determined by densitometry in comparison to known amounts of hnRNP. For electron microscopy of *in vitro* aggregation reactions, samples (10 μ l) were adsorbed onto glow-discharged 300-mesh Formvar/carboncoated copper grid (Electron Microscopy Sciences) and stained with 2% (w/v) aqueous uranyl acetate. Excess liquid was removed, and grids were allowed to air dry. Samples were viewed by a JEOL 1010 transmission electron microscope.

Fly stocks and culture. The wild-type hnRNA2 and mutant hnRNA2(D290V) cDNAs were subcloned into the pUASTattB plasmid, using restriction sites EcoRI and XhoI, creating pUASTattB-wild type hnRNA2 or pUASTattB-mutant hnRNA2(D290V). Wild-type HRB98DE and mutant HRB98DE(D302V) were subcloned into pUASTattB plasmid using NotI and XhoI sites. Flies carrying pUASTattB transgenes were generated by a standard injection and ϕ C31 integrase-mediated transgenesis technique. To express a transgene in muscles, *Mhc-Gal4* was used (from G. Marqués). All *Drosophila* stocks were maintained in a 25 °C incubator with a 12-h day/night cycle.

Adult fly muscle preparation and immunohistochemistry. Adult flies were embedded in a drop of OCT compound (Sakura Finetek) on a slide glass, frozen with liquid nitrogen and bisected sagittally by a razor blade. After fixing with 4% paraformaldehyde in PBS, hemithoraces were stained by Texas Red-X phalloidin (Invitrogen) and DAPI according to manufacturer's instructions. Stained hemithoraces were mounted in 80% glycerol, and the musculature was examined by DMIRE2 (Leica, 10 \times). For hnRNA2 staining, hemithoraces were permeabilized with PBS containing 0.2% Triton X-100 and stained with anti-hnRNA2B1 (EF-67) antibody (Santa Cruz Biotechnology) and Alexa-488-conjugated secondary antibody (Invitrogen). Stained muscle fibres were dissected and mounted in Fluoromount-G (Southern Biotech) and imaged with a Marianas confocal microscope (Zeiss, \times 63).

Fly western blotting. Thoraces of adult flies were prepared and ground in PBS containing 0.2% Triton X-100. After adding SDS sample buffer, samples were boiled for 5 min and analysed by the standard western blotting method provided by Odyssey system (LI-COR) with 4–12% NuPAGE Bis-Tris Gel (Invitrogen) and anti-hnRNA2B1 antibody (Santa Cruz, 1:1,000).

25. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
26. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
27. Goode, M. G. & Rodrigo, A. G. SQUINT: a multiple alignment program and editor. *Bioinformatics* **23**, 1553–1555 (2007).
28. Zwickl, D. J. *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion*. PhD thesis, Univ. Texas at Austin (2006).
29. Johnson, B. S., McCaffery, J. M., Lindquist, S. & Gitler, A. D. A yeast TDP-43 proteinopathy model: exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proc. Natl Acad. Sci. USA* **105**, 6439–6444 (2008).
30. Johnson, B. S. *et al.* TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J. Biol. Chem.* **284**, 20329–20339 (2009).
31. Sun, Z. *et al.* Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol.* **9**, e1000614 (2011).
32. Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway® cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**, 913–919 (2007).
33. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA* **102**, 12825–12830 (2005).
34. Song, Y. *et al.* Role of Hsp70 chaperone in *Saccharomyces cerevisiae* prion seed replication. *Eukaryot. Cell* **4**, 289–297 (2005).

35. Ross, C. D., McCarty, B. M., Hamilton, M., Ben-Hur, A. & Ross, E. D. A promiscuous prion: efficient induction of [URE3] prion formation by heterologous prion domains. *Genetics* **183**, 929–940 (2009).
36. Guthrie, C. & Fink, G. R. *Methods in Ezymology: Guide to Yeast Genetics and Molecular and Cell Biology* 169 (Academic, 2002).
37. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168 (1983).
38. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA* **102**, 12825–12830 (2005).

CLP1 links tRNA metabolism to progressive motor–neuron loss

Toshikatsu Hanada^{1*}, Stefan Weitzer^{1*}, Barbara Mair¹, Christian Bernreuther², Brian J. Wainger^{3,4}, Justin Ichida⁵, Reiko Hanada¹, Michael Orthofer¹, Shane J. Cronin³, Vukoslav Komnenovic¹, Adi Minis⁶, Fuminori Sato⁷, Hiromitsu Mimata⁷, Akihiko Yoshimura⁸, Ido Tamir⁹, Johannes Rainer¹⁰, Reinhard Kofler¹⁰, Avraham Yaron⁶, Kevin C. Eggan⁵, Clifford J. Woolf^{3,11}, Markus Glatzel², Ruth Herbst¹², Javier Martinez¹ & Josef M. Penninger¹

CLP1 was the first mammalian RNA kinase to be identified. However, determining its *in vivo* function has been elusive. Here we generated kinase-dead *Clp1* (*Clp1*^{K/K}) mice that show a progressive loss of spinal motor neurons associated with axonal degeneration in the peripheral nerves and denervation of neuromuscular junctions, resulting in impaired motor function, muscle weakness, paralysis and fatal respiratory failure. Transgenic rescue experiments show that CLP1 functions in motor neurons. Mechanistically, loss of CLP1 activity results in accumulation of a novel set of small RNA fragments, derived from aberrant processing of tyrosine pre-transfer RNA. These tRNA fragments sensitize cells to oxidative-stress-induced p53 (also known as TRP53) activation and p53-dependent cell death. Genetic inactivation of p53 rescues *Clp1*^{K/K} mice from the motor neuron loss, muscle denervation and respiratory failure. Our experiments uncover a mechanistic link between tRNA processing, formation of a new RNA species and progressive loss of lower motor neurons regulated by p53.

RNA molecules undergo co- and post-transcriptional processing, leading to mature, functional RNAs. In mammals and archaea CLP1 proteins are kinases that phosphorylate the 5' hydroxyl ends of RNA^{1–3}. Human CLP1 is a component of the messenger RNA 3'-end cleavage and polyadenylation machinery^{4,5}, and studies in yeast have postulated a function for CLP1 in coupling mRNA 3'-end processing with RNA polymerase II (Pol II) transcriptional termination^{6–8}. Unique to mammals is the association of CLP1 with the tRNA splicing endonuclease (TSEN) complex⁹. TSEN proteins remove the intron present within the anticodon loop of numerous pre-transfer RNAs (pre-tRNAs), generating 5' and 3' tRNA exon halves¹⁰. Within the TSEN complex, CLP1 phosphorylates 3' tRNA exons *in vitro*¹, potentially contributing to tRNA splicing in mammals¹¹. Although CLP1 may participate in multiple RNA pathways and was the first mammalian kinase identified that phosphorylates RNA, the *in vivo* function of CLP1 in mammalian cells has remained elusive.

We report the generation and phenotypic analysis of CLP1 kinase-dead mice. These mice develop a progressive loss of lower motor neurons, resulting in fatal deterioration of motor function. We also show that inactivation of CLP1 kinase activity results in the accumulation of previously unreported tyrosine tRNA fragments that sensitize cells to activation of p53 in response to oxidative stress.

Neonatal lethality of *Clp1*^{K/K} mice

To assess the *in vivo* function of CLP1, first we generated global *Clp1*-knockout mice. We never obtained any viable *Clp1*-null offspring, even when analysed at embryonic day (E)6.5, indicating very early embryonic lethality. Consequently, we generated mice carrying a

single amino-acid change, lysine to alanine at position 127 (K127A), which is located within the Walker A ATP-binding motif (Supplementary Fig. 1a, b). This mutation abolishes CLP1 kinase activity¹. Mice heterozygous for the K127A substitution (*Clp1*^{K/+}) were intercrossed to generate homozygous offspring (*Clp1*^{K/K}). Western blotting showed that the CLP1 K127A mutant protein was expressed at normal levels, and that the *Clp1*^{K/K} mutation impaired 5' phosphorylation of a small duplex RNA substrate (Supplementary Fig. 1c, d). Therefore, we successfully generated a knock-in mouse expressing kinase-dead CLP1.

Clp1^{K/K} mice were born at a normal Mendelian ratio. However, on a C57BL/6 background, all *Clp1*^{K/K} mice died within hours of birth, probably owing to respiratory failure (Fig. 1a, b). This phenotype had complete penetrance ($n > 50$). Embryos and newborn *Clp1*^{K/K} mice had overtly normal lung development and morphogenesis, as indicated by caveolin 1, surfactant A and surfactant C expression (Supplementary Fig. 2). However, all newborn *Clp1*^{K/K} mice and E18.5 embryos exhibited a lordotic body posture and dropping forelimbs, indicative of impaired motor functions (Fig. 1a and Supplementary Fig. 3a). Newborn *Clp1*^{K/K} mice also showed reduced birth weight and were hyporesponsive to stimuli (Supplementary Fig. 3b); similar phenotypes in KIF1B-mutant mice have been ascribed to motorsensory neuronal defects¹². We therefore analysed neuromuscular junctions (NMJs) in the diaphragm.

Control E18.5 embryos showed the characteristic innervation pattern of the phrenic motor nerve bundle and had normal NMJs, defined by colocalization of presynaptic terminals with postsynaptic clusters of acetylcholine receptors (AChRs) as well as the presence of

¹IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna 1030, Austria. ²Institute of Neuropathology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany. ³Program in Neurobiology and F. M. Kirby Neurobiology Center, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ⁴Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵Department of Stem Cell and Regenerative Biology, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. ⁶Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel. ⁷Department of Urology, Oita University Faculty of Medicine, 1-1 Idaigaoka, Hasama-machi, Yufu, Oita 879-5593, Japan. ⁸Department of Microbiology and Immunology, Keio University School of Medicine, Shinjuku-ku, Tokyo 160-8582, Japan. ⁹Campus Science Support Facilities GmbH, Dr. Bohr-Gasse 3, Vienna 1030, Austria. ¹⁰Division Molecular Pathophysiology, Biocenter, Medical University of Innsbruck, Innsbruck 6020, Austria. ¹¹Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹²Center for Brain Research, Medical University of Vienna, Spitalgasse 4, Vienna 1090, Austria.

*These authors contributed equally to this work.

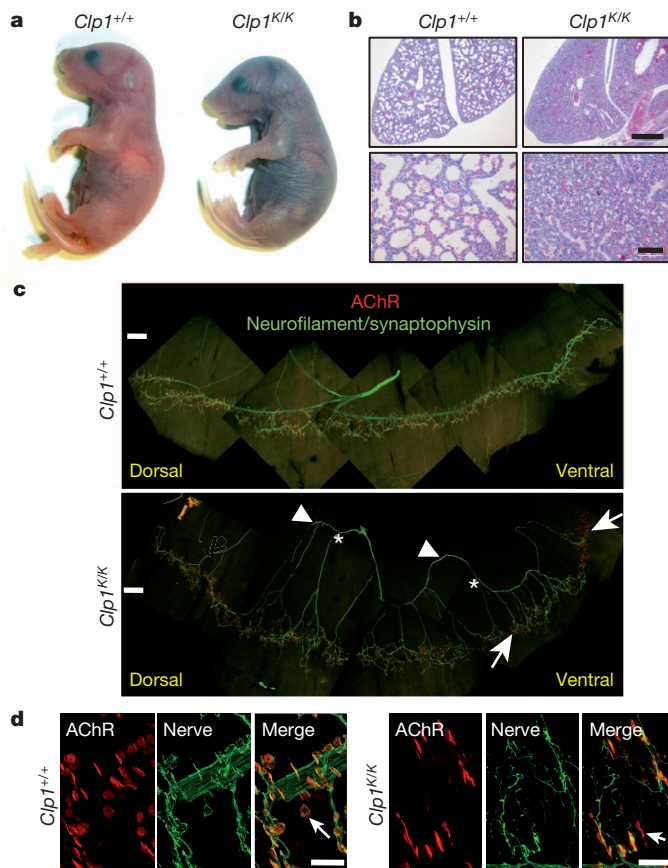


Figure 1 | Respiratory failure and impaired innervation of the diaphragm. **a, b,** Appearance (**a**) and lung histology (**b**; haematoxylin and eosin) of newborn $Clp1^{+/+}$ and $Clp1^{K/K}$ littermates on a C57BL/6 background. Scale bars: top, 500 μ m; bottom, 100 μ m. **c, d,** Whole-mount immunostaining of diaphragm muscle of E18.5 $Clp1^{+/+}$ and $Clp1^{K/K}$ littermates showing postsynaptic AChR clusters (red, α -bungarotoxin), and innervating motor axons and presynaptic nerve terminals (green, neurofilament/synaptophysin immunostaining). Asterisks indicate defasciculated main axons mislocalized to the periphery of the muscle; arrowheads indicate secondary branches coming from the main axon; arrows indicate areas of denervation. Magnifications in **d** indicate severely disturbed NMJs in E18.5 $Clp1^{K/K}$ embryos (arrow). Scale bars: **c**, 250 μ m; **d**, 25 μ m.

S100⁺ Schwann cells in the endplate (Fig. 1c and Supplementary Fig. 4). E18.5 $Clp1^{K/K}$ embryos showed defasciculation of the main phrenic nerve bundle; primary branches were mislocalized to the periphery, and denervation of the ventral and dorsal diaphragm was prominent in all $Clp1^{K/K}$ mutants (Fig. 1c). NMJs were formed, but axon terminals appeared undifferentiated, with smaller AChR clusters (Fig. 1d). S100 expression at the NMJ was absent in $Clp1^{K/K}$ embryos, although the Schwann cells seemed to be functionally intact as surviving peripheral axons were myelinated (Supplementary Fig. 4). The development and morphology of the heart, liver, kidney, colon, bladder, spleen and thymus appeared normal at E18.5. Thus, all newborn $Clp1^{K/K}$ mice show impaired innervation of the diaphragm, which seems to cause lethal respiratory failure and neonatal death.

Embryonic loss of motor neurons

We next assessed NMJs in the diaphragm during embryogenesis. In $Clp1^{K/K}$ embryos, denervation was not found at E14.5. However, we observed partial denervation and pronounced alteration in NMJ morphology at E16.5, followed by a severe defect in the innervation of the NMJs of the diaphragm at E18.5 (Fig. 1c, d and Supplementary Figs 4–6). Moreover, whereas $Clp1^{K/K}$ embryos had normal numbers of choline acetyl transferase (ChAT)-expressing spinal motor neurons

at E14.5 and E16.5, the numbers of ChAT⁺ motor neurons markedly declined in the spinal cord of E18.5 $Clp1^{K/K}$ embryos (Fig. 2a, b and Supplementary Fig. 7a, b). The numbers of motor neurons also declined in wild-type embryos due to pruning, which has been linked to oxidative stress exposure¹³. NeuN staining to detect all neurons showed that the observed reduction in neuronal numbers was due to the loss of ChAT⁺ motor neurons (Supplementary Fig. 7c).

To confirm loss of motor neurons in the spinal cord, we crossed green fluorescent protein (GFP)-tagged Hb9 (also known as Mn1) transgenic mice onto a $Clp1^{K/K}$ background. Genetic GFP tagging showed that E14.5 $Clp1^{K/K}$ embryos have similar numbers of Hb9⁺ cells as age-matched control embryos, but the numbers of Hb9-GFP⁺ cells declined in E18.5 $Clp1^{K/K}$ embryos (Supplementary Fig. 7d). In whole-mount visualizations of Hb9-GFP⁺ E10.5 embryos, both development and segmental-motor-axon outgrowth were comparable between $Clp1^{K/K}$ and control embryos; three-dimensional reconstructions further demonstrated comparable motor-axon outgrowth in the thoracic region (segment T6–T7) in E12.5 embryos

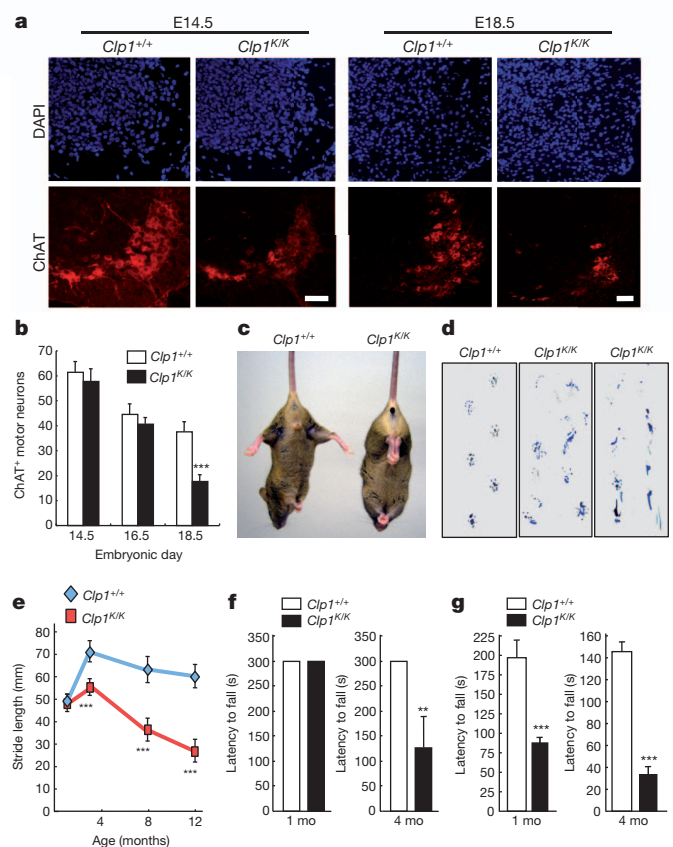


Figure 2 | Viable $Clp1^{K/K}$ embryos develop neuromuscular atrophy.

a, Immunostaining for ChAT⁺ (red) motor neurons in the cervical (C4–C5) spinal cord of E14.5 and E18.5 $Clp1^{+/+}$ and $Clp1^{K/K}$ littermates on a C57BL/6 background. 4',6-Diamidino-2-phenylindole (DAPI) staining is shown. Scale bar: 50 μ m. **b,** Mean numbers \pm standard deviation (s.d.) of ChAT⁺ motor neurons in the cervical (C4–C5) spinal cord of E14.5, E16.5 and E18.5 $Clp1^{+/+}$ and $Clp1^{K/K}$ littermates. $n = 5$ mice per group. *** $P < 0.001$ (t -test). **c,** Muscle weakness (impaired spreading of hind legs) in 3-month-old $Clp1^{K/K}$ mice. **d, e,** Impaired walking strides in 1-, 3-, 8- and 12-month-old $Clp1^{K/K}$ mice compared to $Clp1^{+/+}$ littermates. Representative images in **d** are from 12-month-old mice, demonstrating shortened walking strides and paralysis in $Clp1^{K/K}$ mice. Data in **e** are mean values \pm s.d. $n = 5$ mice per group. *** $P < 0.001$ (t -test). **f, g,** Progressive impairment of motor functions in viable $Clp1^{K/K}$ mice as assessed by the latency to fall in a fixed (**f**) and accelerated (**g**) Rotarod test. Data (mean values \pm s.d.) are from 1-month-old (mo) $Clp1^{+/+}$ ($n = 8$) and $Clp1^{K/K}$ ($n = 7$) mice, and 4-month-old $Clp1^{+/+}$ ($n = 16$) and $Clp1^{K/K}$ ($n = 14$) mice. ** $P < 0.01$, *** $P < 0.001$ (t -test).

(Supplementary Fig. 8a, b). Moreover, the overall morphology of motor terminal branches and the pathfinding of motor neurons at the intercostal muscles was comparable (Supplementary Fig. 8c). Thus, genetic inactivation of the RNA kinase function of CLP1 results in progressive loss of motor neurons in the cervical and lumbar spinal cord, with aberrant innervations and formation of NMJs in the diaphragm, and neonatal death.

Progressive loss of motor functions

The 100% lethality of *Clp1*^{K/K} pups at birth was observed on a C57BL/6 background. Similarly, we observed 100% lethality on a BALB/c background. However, when crossing the mutation onto a CBA/J mouse background, we obtained viable pups homozygous for the *Clp1*^{K/K} mutation (Supplementary Fig. 9a). These mice showed motor ataxia (Supplementary Fig. 9b), impaired muscle strength, an altered walking stride, and diminished balance as determined by the fixed and accelerated Rotarod tests (Fig. 2c–g), all indications of motor defects. The muscle weakness and impaired motor functions observed in these mice were progressive (Fig. 2e–g). Mice started to die around week 23 after birth, and older *Clp1*^{K/K} mice developed limb paralysis (Fig. 2d and Supplementary Fig. 9a, c). Sensory responses to noxious heat, mechanical stimulation and capsaicin-induced pain appeared normal in *Clp1*^{K/K} littermates (Supplementary Fig. 10).

Whereas *Clp1*^{K/K} neonates on the CBA/J background had apparently normal numbers of spinal motor neurons, 4-month-old *Clp1*^{K/K} mice showed both a marked reduction in ChAT⁺ and, as a second marker, SMI-32⁺ spinal motor neurons (Fig. 3a, b and Supplementary Fig. 11). The effects of the *Clp1* mutation on upper motor neurons need to be assessed. Consistent with lower motor neuron loss, we observed axonopathy of peripheral nerves such as the sciatic nerve (Fig. 3c, d and Supplementary Fig. 12a–c). Immunostaining showed amyloid- β precursor protein (APP)-positive structures, a marker for axonal injury¹⁴, in the sciatic nerves of adult *Clp1*^{K/K} mice (Fig. 3e). In line with motor-neuron loss, we observed a marked reduction in the number of large diameter (type 1A alpha) fibres, whereas the smaller sensory fibres were largely preserved (Supplementary Fig. 12d). Dorsal root ganglion (DRG) sensory neurons showed normal morphology and outgrowth. Whether developmental alterations in myelination also contribute to the phenotype needs to be further assessed. We also observed regional denervation and fragmentation of NMJs in the diaphragm, and in various limb and head skeletal muscles, resulting in skeletal muscle atrophy; slow-twitch muscles (soleus, gluteus) were less affected than the fast-twitch extensor digitorum longus or gastrocnemius muscles (Fig. 3f, g and Supplementary Figs 13–15). Similar results were obtained in SOD1 transgenic mice¹⁵. Thus, *Clp1*^{K/K} mice develop progressive loss of spinal motor neurons and exhibit defective NMJs, resulting in impaired motor functions, muscular atrophy and limb paralysis.

CLP1 promotes efficient tRNA exon generation

To assess potential roles in RNA metabolic processes, first we evaluated the function of CLP1 in mRNA 3'-end cleavage⁴. The kinetics of a pre-mRNA cleavage reaction in nuclear extracts from wild-type and *Clp1*^{K/K} mouse embryonic fibroblasts (MEFs) were similar (Supplementary Fig. 16a), suggesting that ATP binding and/or hydrolysis by CLP1 are dispensable for mRNA 3'-end cleavage. CLP1 has been implicated in RNA interference (RNAi) by its ability to phosphorylate small interfering RNAs *in vitro*¹. We therefore assessed whether cellular microRNA (miRNA) processing or stability were affected by the CLP1 K127A mutation. However, the levels of various pre- and mature miRNAs did not differ between wild-type and *Clp1*^{K/K} MEFs and tissues, and no effect on miRNA function was detected using a luciferase reporter assay (Supplementary Fig. 16b, c). Furthermore, we found no significant differences in pre-mRNA splicing (Supplementary Fig. 17).

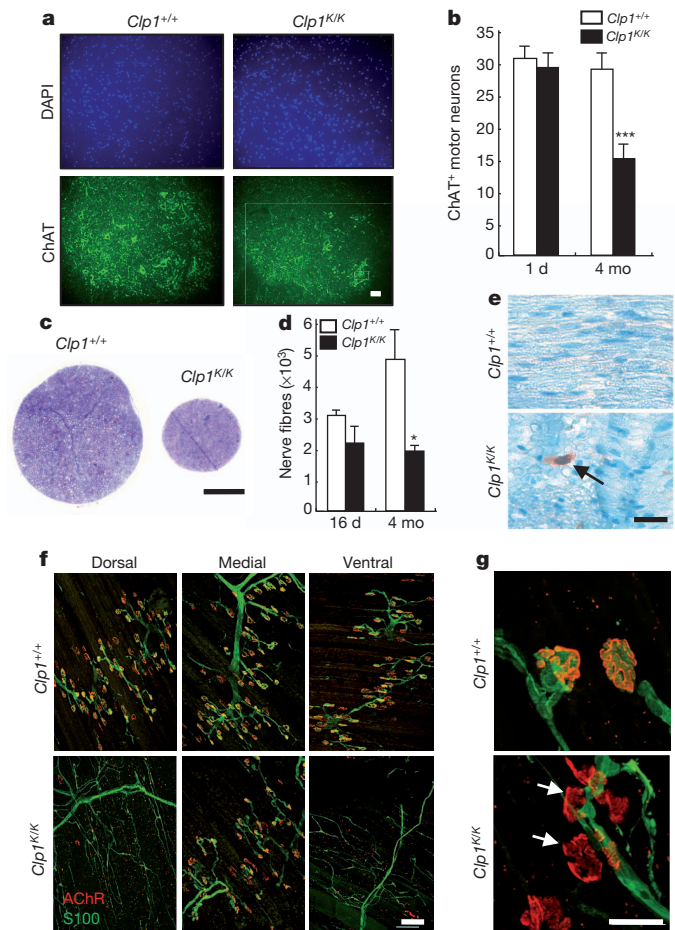


Figure 3 | Progressive loss of lower motor neurons. **a**, Immunostaining for ChAT⁺ (green) motor neurons in the lumbar (L5) spinal cord of 4-month-old *Clp1*^{+/+} and *Clp1*^{K/K} littermates on a CBA/J background. DAPI counterstaining. **b**, Mean numbers \pm s.d. of ChAT⁺ motor neurons in the lumbar (L5) spinal cord of 1-day (d) and 4-month-old (mo) *Clp1*^{+/+} and *Clp1*^{K/K} littermates. $n = 5$ mice per group. *** $P < 0.001$ (t -test). **c**, Semi-thin cross-section of the sciatic nerve of 4-month-old *Clp1*^{+/+} and *Clp1*^{K/K} littermates. Toluidin blue staining. **d**, Mean numbers \pm s.d. of total nerve fibres in the sciatic nerves of 16-day and 4-month-old *Clp1*^{+/+} and *Clp1*^{K/K} littermates. $n = 3$ mice per group. * $P < 0.05$ (t -test). **e**, Immunohistochemical detection of APP (arrow) in the sciatic nerve of a 12-month-old *Clp1*^{K/K} mouse. **f**, **g**, Whole-mount immunostaining depicting NMJs in the diaphragms of 5-month-old *Clp1*^{+/+} and *Clp1*^{K/K} littermates. Postsynaptic AChR clusters are stained with α -bungarotoxin (red) and Schwann cells are labelled with anti-S100 antibodies (green). Arrows in **g** show NMJ fragmentations. Scale bars: **a**, 50 μ m; **b**, 200 μ m; **c**, 100 μ m; **d**, 25 μ m; **g**, 100 μ m.

It has been reported that CLP1 is part of the human TSEN complex⁹ and is able to phosphorylate tRNA 3' exons¹. We therefore tested whether the *Clp1*^{K/K} mutation affects tRNA splicing. Notably, nuclear extracts from *Clp1*^{K/K} MEFs generated significantly lower levels of tRNA exons (Fig. 4a). In addition, tRNA 3' exon halves formed by incubation with nuclear extracts isolated from *Clp1*^{K/K} MEFs lacked a 5' phosphate group; 5' phosphorylation and exon generation were rescued by ectopic expression of wild-type CLP1 in *Clp1*^{K/K} MEFs (Supplementary Fig. 18a–e). Similarly, pre-tRNA cleavage was partially restored in *Clp1*^{K/K} MEFs stably expressing wild-type CLP1, whereas overexpression of the CLP1 K127A mutant inhibited RNA phosphorylation (Supplementary Fig. 19a–d). Affinity-purified CLP1(K127A)-containing TSEN complexes were deficient in the generation of tRNA exons, most probably due to reduced levels of TSEN2, TSEN34 and TSEN54 subunits (Fig. 4b, c); mRNA levels of TSEN2, TSEN34 and TSEN54 were not affected in *Clp1*^{K/K} MEFs (data not shown). Thus, the RNA kinase activity of CLP1 is important

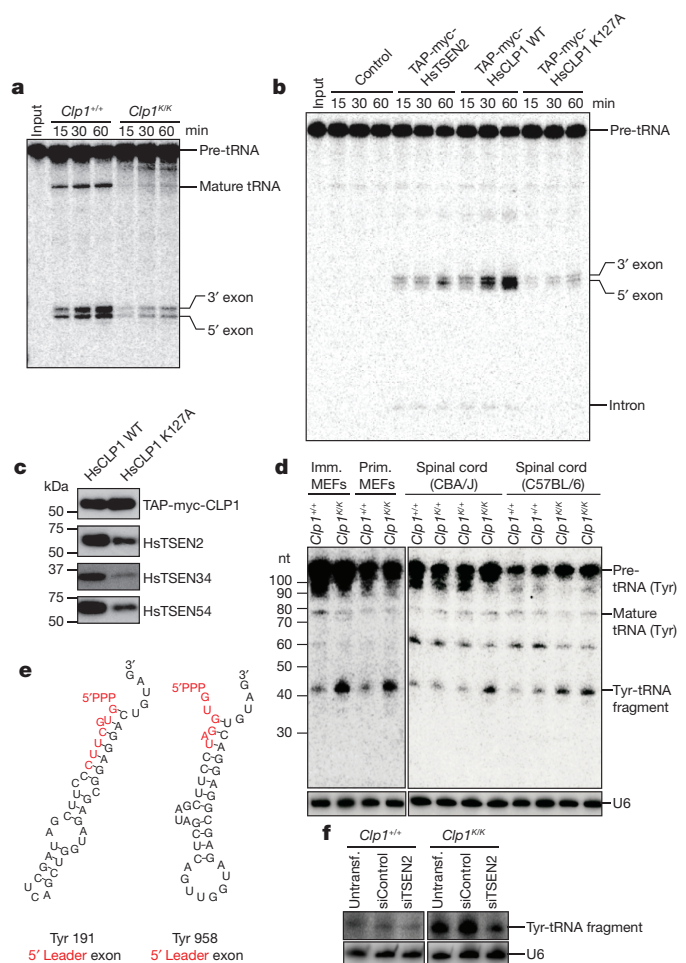


Figure 4 | Identification of a novel tRNA fragment. **a**, The RNA kinase activity of CLP1 is required for efficient tRNA exon generation *in vitro*. An internally labelled intron-containing yeast pre-tRNA^{Phe} was incubated with nuclear extracts from immortalized *Clp1*^{+/+} and *Clp1*^{K/K} MEFs. Pre-tRNA processing was monitored by electrophoresis. **b**, One-step purified human (Hs) TAP-myc-HsCLP1 (wild-type (WT) or K127A) and TAP-myc-HsTSEN2 complexes were assayed for pre-tRNA cleavage after elution with tobacco etch virus (TEV) protease (TEV eluate) by incubation with an internally labelled pre-tRNA^{Phe}. Eluates from HeLa cells expressing the vector without insert were used as control. **c**, The level of TSEN proteins in TAP-myc-HsCLP1 (wild type and K127A) TEV eluates was examined by western blot. Similar CLP1 protein levels were confirmed using anti-myc antibodies. **d**, ~41–46-nucleotide (nt) Tyr-tRNA fragments accumulate in primary (Prim.) and immortalized (Imm.) *Clp1*^{K/K} MEFs and spinal cords of 4-month-old CBA/J *Clp1*^{K/K} and newborn C57BL/6 *Clp1*^{K/K} mice. Northern blot analyses of total RNA were performed with a DNA/LNA probe complementary to the 5' exon of Tyr-tRNA. U6 RNA served as a loading control. **e**, Secondary structure of the two most abundant tRNA fragment species, Tyr-tRNA *M. musculus* chr14.tRNA191-TyrGTA and *M. musculus* chr13.tRNA958-TyrGTA (mouse July 2007 (mm9) genome assembly), as predicted by RNAfold (Vienna RNA Secondary Structure Package). **f**, Detection of Tyr-tRNA fragments from *Clp1*^{+/+} and *Clp1*^{K/K} MEFs left untransfected (Untransf.) or transfected with control siRNAs or siRNAs against TSEN2.

for the integrity of the TSEN complex and for efficient generation of tRNA exons (Supplementary Fig. 20).

Accumulation of novel tyrosine tRNA fragments

Whereas most tRNAs do not contain introns, all mouse tyrosine (Tyr)-tRNA genes contain an intronic sequence. Notably, in *Clp1*^{K/K} MEFs we detected accumulation of ~41–46-nucleotide Tyr-tRNA fragments using a northern probe against the 5' exon (Fig. 4d). RNA sequencing revealed that this fragment comprises a 5' leader followed by 5' exon

Tyr-tRNA sequences (Fig. 4e and Supplementary Fig. 21a, b). Enzymatic treatments uncovered a 5' triphosphate modification (Supplementary Fig. 21c), indicating that this novel tRNA fragment contains a full-length 5' leader sequence starting with the transcription initiator PPP-nucleotide. RNAi-mediated silencing of TSEN2 in MEFs resulted in decreased levels of tRNA fragments (Fig. 4f and Supplementary Fig. 21d). Bioinformatic and northern blotting analyses revealed minor accumulation of such 5' fragments from other intron-containing tRNAs, mostly arginine tRNAs (Supplementary Fig. 22).

A similar accumulation of tyrosine tRNA fragments was observed in the spinal cord of C57BL/6 *Clp1*^{K/K} neonates and 4-month-old *Clp1*^{K/K} mice on a CBA/J background (Fig. 4d). Moreover, we observed increased levels of these tRNA fragments in cortex, muscle, heart, kidney, muscle and liver (Supplementary Fig. 23a). Steady-state levels of mature tRNAs seemed to be normal in MEFs, the spinal cord of C57BL/6 *Clp1*^{K/K} neonates and in postnatal CBA/J *Clp1*^{K/K} mice (Supplementary Fig. 23b). Thus, loss of CLP1 kinase activity results in the accumulation of novel RNA fragments derived from pre-tRNA.

Oxidative-stress-induced cell death

Other tRNA fragments, for example, tRNA-derived, stress-induced small RNAs (tiRNAs), are generated in the cytoplasm by the endonuclease angiogenin acting on mature tRNAs^{16,17}. Similar to tiRNAs, Tyr-tRNA fragments were also induced by H₂O₂, but were not generated by angiogenin. Rather, Tyr-tRNA fragments were present in the nucleus and were derived from pre-tRNAs actively transcribed by RNA Pol III (Supplementary Fig. 24a–d). Because Tyr-tRNA fragments were strongly induced by H₂O₂, we proposed that CLP1 might have a role in the oxidative stress response. In wild-type MEFs, the Tyr-tRNA fragments also accumulated after exposure to the reactive oxygen species (ROS) inducers glucose oxidase, menadione and paraquat dichloride, but not nefazodone (Supplementary Fig. 25a–d). Treatment with the mitochondrial 'poisons' carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP) and rotenone, the apoptosis inducer staurosporine, the protein-synthesis inhibitor puromycin, and the DNA-damaging agent camptothecin did not trigger the accumulation of Tyr-tRNA fragments (Supplementary Fig. 25e, f). Importantly, on H₂O₂ and glucose oxidase challenge, we observed increased death of *Clp1*^{K/K} MEFs; re-expression of wild-type CLP1 restored the survival rate to that of control MEFs (Fig. 5a and Supplementary Fig. 26a).

Extending our studies to neurons, we trans-differentiated wild-type and *Clp1*^{K/K} MEFs into Hb9-GFP⁺ motor neurons¹⁸, and observed typical sodium and potassium currents, as well as normal action potentials and responses to excitatory and inhibitory transmitters (Fig. 5b and Supplementary Fig. 26b–f). However, the resting membrane potential of *Clp1*^{K/K} motor neurons was depolarized by over 10 mV relative to control motor neurons (-57.0 ± 2.0 mV for $n = 11$ *Clp1*^{+/+} and -46.9 ± 3.0 mV for $n = 10$ *Clp1*^{K/K} motor neurons; t -test < 0.01), suggesting that *Clp1*^{K/K} motor neurons exhibit ion-exchange abnormalities. Cell input resistance was similar between the two groups of neurons (367 ± 94 MΩ for $n = 11$ *Clp1*^{+/+} and 511 ± 63 MΩ for $n = 10$ *Clp1*^{K/K} motor neurons; t -test > 0.2). Importantly, similar to MEFs, trans-differentiated *Clp1*^{K/K} motor neurons were more sensitive to H₂O₂-induced cell death than wild-type motor neurons (Fig. 5c). Thus, loss of the catalytic activity of CLP1 results in enhanced motor-neuron death in response to oxidative stress.

Motor-neuron loss is mediated by p53

Oxidative stress has been linked to a p53-regulated cell death pathway through serine-18 phosphorylation, which has been shown to regulate p53-dependent transcription^{19–21}. In response to H₂O₂, *Clp1*^{K/K} MEFs showed hyperphosphorylation of p53 at serine 18 and increased induction of the p53-target gene p21 (also known as CDKN1A); re-expression of wild-type CLP1 rescued the H₂O₂-induced serine-18

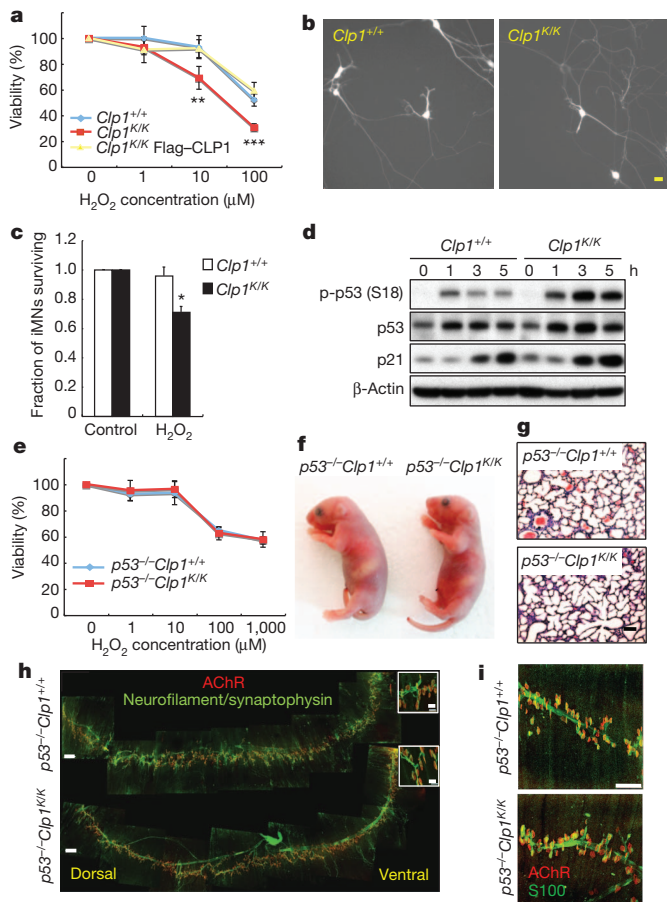


Figure 5 | Neonatal lethality and motor-neuron loss are mediated by p53. **a**, Viability of *Clp1*^{+/+}, *Clp1*^{K/K} and *Clp1*^{K/K} 3T3 MEFs rescued with Flag-tagged wild-type CLP1. Cells were exposed in triplicate to the indicated concentrations of H₂O₂ for 1 h, and mean viability was determined 12 h later. Data represent the means \pm s.d. of three independent experiments. ***P* < 0.01, ****P* < 0.001 (analysis of variance (ANOVA)). **b**, Representative images of Hb9-GFP⁺ trans-differentiated *Clp1*^{+/+} and *Clp1*^{K/K} motor neurons. Scale bar: 20 μ m. **c**, Wild-type and *Clp1*^{K/K} trans-differentiated motor neurons (iMNs) were treated in duplicate cultures with 100 μ M H₂O₂ or water (control) for 3 h. Mean survival \pm s.d. was scored after 48 h. **P* < 0.05 (paired *t*-test). **d**, p53 serine-18 phosphorylation (p) in H₂O₂-treated (100 μ M for 1 h) *Clp1*^{K/K} and *Clp1*^{+/+} primary MEFs. Numbers indicate hours of recovery after stimulation. Time point 0 indicates no stimulation. β -Actin levels served as loading control. **e**, Mean viability \pm s.d. of *p53*^{-/-} *Clp1*^{+/+} and *p53*^{-/-} *Clp1*^{K/K} MEFs challenged in triplicate with the indicated concentrations of H₂O₂ for 1 h. Cell viability was determined 12 h later. Data represent the means \pm s.d. of three independent experiments. **P* < 0.05 (*t*-test). **f, g**, General appearance (**f**) and lung structures (**g**; haematoxylin and eosin) of newborn *p53*^{-/-} *Clp1*^{+/+} and *p53*^{-/-} *Clp1*^{K/K} littermates on a C57BL/6 background. Scale bar: 50 μ m. **h**, Whole-mount immunostaining of the diaphragms of 3-day-old *p53*^{-/-} *Clp1*^{+/+} and *p53*^{-/-} *Clp1*^{K/K} littermates showing postsynaptic AChR clusters (red; α -bungarotoxin) and axons and branched presynaptic terminals of the phrenic nerve (green; neurofilament/synaptophysin immunostaining). Scale bar: 250 μ m; inset: 25 μ m. **i**, Whole-mount immunostaining of diaphragms of E18.5 *p53*^{-/-} *Clp1*^{+/+} and *p53*^{-/-} *Clp1*^{K/K} littermates showing postsynaptic AChR clusters (red; α -bungarotoxin) and Schwann cells (green; S100). Scale bar: 75 μ m.

hyperphosphorylation (Fig. 5d and Supplementary Fig. 27a, b). Similarly, we found p53 serine-18 hyperphosphorylation in *Clp1*^{K/K} MEFs in response to the ROS-inducing agent glucose oxidase but not the DNA-damaging agent camptothecin (Supplementary Fig. 27c, d). Most importantly, genetic inactivation of p53 rescued *Clp1*^{K/K} MEFs from enhanced death after H₂O₂ treatment (Fig. 5e). Overexpression of two different tyrosine tRNA fragments in the mouse motor neuron cell line NSC-34 also resulted in enhanced p53 activation in response

to H₂O₂ (Supplementary Fig. 28a, b). Thus, kinase-dead CLP1 renders MEFs more susceptible to oxidative-stress-induced cell death through a p53-regulated pathway.

Notably, we observed a complete rescue of neonatal lethality in *p53*^{-/-} *Clp1*^{K/K} mice on the 100% lethal C57BL/6 background (Fig. 5f). The viable *p53*^{-/-} *Clp1*^{K/K} pups showed normal extension of the lungs (Fig. 5g), normal innervation of the diaphragm, and rescue of NMJ formation (Fig. 5h, i). The numbers of ChAT⁺ (Supplementary Fig. 29a, b) and SMI-32-labelled (data not shown) motor neurons in the spinal cord were comparable to that of newborn and 1-month-old control mice. Furthermore, 1-month-old *p53*^{-/-} *Clp1*^{K/K} mice did not show muscle weakness (Supplementary Fig. 29c–e). Older *p53*^{-/-} *Clp1*^{K/K} mice could not be assessed because they developed tumours due to the loss of p53. To test whether oxidative stress is involved in the phenotype, we treated pregnant *Clp1*^{K/K} females (crossed to *Clp1*^{+/+} males on the lethal C57BL/6 background) with the ROS scavenger *N*-acetylcysteine (NAC). NAC treatment resulted in viable *Clp1*^{K/K} pups and partially restored innervation of the diaphragm (Supplementary Fig. 30). However, pups from NAC-treated mothers died within a week after birth, probably because *in vivo* ROS scavenging by NAC is incomplete and/or other pathways contribute to the phenotype. Thus, ROS and p53 constitute critical *in vivo* pathways that mediate motor neuronal loss and neonatal death of *Clp1*^{K/K} mutant mice.

Transgenic rescue of motor-neuron defects

To provide definitive proof that the CLP1 kinase-dead mutation acts in motor neurons, we introduced a Flag-tagged wild-type CLP1-IRES-eGFP transgene under the control of the motor-neuron promoter Hb9 (ref. 22) into control and *Clp1*^{K/K} mice on a C57BL/6 background (Fig. 6a and Supplementary Fig. 31a, b). Transgenic expression of wild-type CLP1 completely rescued neonatal lethality of *Clp1*^{K/K} mice and restored normal numbers of ChAT⁺ motor neurons in the spinal cord of newborns (Fig. 6b, c and Supplementary Fig. 31c). Moreover, Hb9-CLP1 transgenic *Clp1*^{K/K} mice did not accumulate Tyr-tRNA fragments in the spinal cord (Fig. 6d), providing direct evidence that accumulation of these tRNA fragments is dependent on the loss of CLP1 activity. Phrenic nerve defasciculation and mislocalization, and the defects in NMJ formation and innervation of the diaphragm in *Clp1*^{K/K} mice were also rescued by transgenic expression of wild-type CLP1 in *Clp1*^{K/K} mice; similar results were obtained for intercostal muscles (Fig. 6e and Supplementary Fig. 32a–c). Moreover, S100⁺ Schwann cells were restored (Fig. 6f), indicating that the observed absence of terminal Schwann cells is probably due to the impaired function of CLP1 in motor neurons.

Specific rescue of CLP1 expression in motor neurons but not other cell types led us to test whether kinase-dead CLP1 might affect general metabolism, which could then contribute to the observed phenotype. Using calorimetric experiments, *Clp1*^{K/K} mice did not exhibit any apparent alterations in food and water intake, O₂ consumption, CO₂ production, respiratory exchange rate or heat generation as compared to the *Clp1*^{+/+} transgenic littermates; moreover, heat generation upon cold exposure and recovery of body temperature, as a measure for sympathetic nerve activity²³, were comparable (Supplementary Fig. 33a–h). These transgenic rescue experiments provide direct evidence that CLP1 acts in lower motor neurons and that mutant-CLP1-mediated motor defects are responsible for the neonatal lethality phenotype.

Conclusions

Our results provide the first report on the *in vivo* function of the RNA kinase CLP1. CLP1 kinase-dead mice develop progressive loss of spinal motor neurons, leading to muscle denervation and paralysis. Inactivation of CLP1 kinase activity results at the same time in poor generation of tRNA exon halves and accumulation of novel, hitherto undescribed 5' leader exon tRNA fragments. This paradoxical

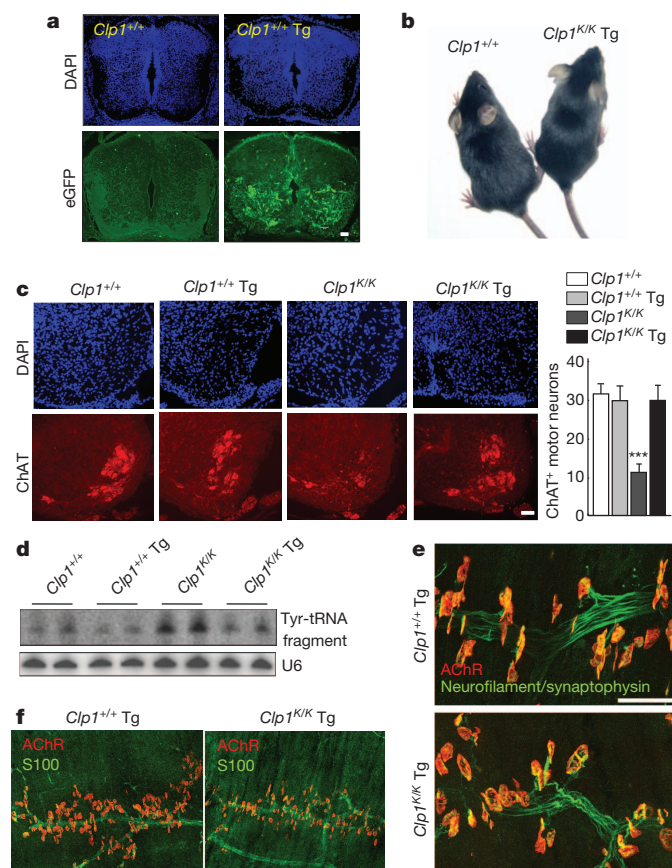


Figure 6 | CLP1 acts in motor neurons. **a**, Cryosections of lumbar (L5) spinal cord from E14.5 *Clp1*^{+/+} and *Clp1*^{+/+} Hb9-CLP1 transgenic (*Clp1*^{+/+} Tg) embryos. Transgene expression was localized by eGFP and anti-eGFP antibodies. DAPI counterstaining is shown. **b**, Appearance of 1-month-old *Clp1*^{+/+} *Clp1*^{K/K} transgenic littermates on a C57BL/6 background. **c**, Immunostaining for ChAT⁺ (red) motor neurons in the lumbar (L5) spinal cord of newborn *Clp1*^{+/+}, *Clp1*^{+/+} transgenic, *Clp1*^{K/K} and *Clp1*^{K/K} transgenic mice on a C57BL/6 background. DAPI staining. Right panel shows mean numbers \pm s.d. of ChAT⁺ motor neurons. $n = 5$ mice per group. *** $P < 0.001$ (ANOVA). **d**, Levels of Tyr-tRNA fragments in the spinal cord of *Clp1*^{+/+}, *Clp1*^{+/+} transgenic, *Clp1*^{K/K} and *Clp1*^{K/K} transgenic mice. Northern blot analyses of total RNA using a probe complementary to the 5' exon of Tyr-tRNA. U6 RNA served as a loading control. **e**, Whole-mount immunostaining depicting NMJs in diaphragms of 3-day-old *Clp1*^{+/+} transgenic and *Clp1*^{K/K} transgenic littermates. Postsynaptic AChR clusters (red; α -bungarotoxin) and innervating motor axons and presynaptic nerve terminals (green; neurofilament/synaptophysin immunostaining) are shown. **f**, Whole-mount immunostaining of the diaphragms of 3-day-old *Clp1*^{+/+} transgenic, *Clp1*^{K/K} and *Clp1*^{K/K} transgenic littermates showing postsynaptic AChR clusters (red; α -bungarotoxin) and S100⁺ Schwann cells (green). Scale bars: **a**, **c**, **e**, 50 μ m; **f**, 100 μ m.

observation could be explained by a defect in tRNA exon ligation in a CLP1 kinase-dead background regulated by oxidative stress. Different tRNA fragments have been previously reported in a variety of organisms^{24,25}. For example, the cytoplasmic pre-tRNA-derived tRF-1001 fragment, a 3' trailer molecule, seems to have a role in cell viability²⁶. Stress-induced tRNAs, derived from the processing of mature tRNAs in the cytoplasm¹⁶, inhibit initiation of protein translation by displacing the eIF4F complex from mRNAs¹⁷. We did not observe translation inhibition by our novel Tyr-tRNA fragments in metabolic labelling experiments; rather, we found that such fragments sensitize cells to oxidative-stress-induced activation of the p53 tumour suppressor pathway. The exact molecular mechanism(s) by which these tRNA fragments couple to the p53 pathway need to be determined. Importantly, motor-neuron loss was rescued *in vivo* either by reducing oxidative stress or by genetically inactivating p53.

Our experiments uncover an unexpected mechanistic link between tRNA processing, formation of a new RNA species, and a p53-regulated progressive loss of lower motor neurons. These results provide a conceptual and experimental framework for how alterations in tRNA metabolism can affect spinal motor neurons, and might help explain fundamental molecular principles in diseases such as amyotrophic lateral sclerosis or spinal muscular atrophy.

METHODS SUMMARY

Generation of kinase-dead *Clp1*^{K/K} and Hb9-CLP1 transgenic mice. Complete mutant and knock-in *Clp1*^{K/K} mice were generated by homologous recombination. The *Clp1*^{K/K} allele was backcrossed five times to C57BL/6 or CBA/J mice. p53-deficient mice and Hb9-GFP transgenic mice were obtained from The Jackson Laboratory. Hb9-CLP1 transgenic mice containing the Hb9 promoter driving Flag-tagged mouse CLP1 and eGFP were generated in house. All mice were maintained according to institutional guidelines.

Phenotyping. Behavioural phenotyping and assessment of motor functions were performed as described in Supplementary Information. Diaphragm and intercostals muscles were stained as described previously²⁷. Anti-neurofilament and anti-synaptophysin antibodies were used to label axons and presynaptic nerve terminals, respectively. Antibodies to S100 were used to visualize Schwann cells. Postsynaptic AChRs were detected with Alexa-594-conjugated α -bungarotoxin. Spinal motor neurons were stained with anti-ChAT and anti-SMI-32 antibodies. MEFs and trans-differentiated motor neurons were generated as described¹⁸.

Pre-tRNA splicing. Pre-tRNA substrates were generated by *in vitro* transcription. Labelled yeast pre-tRNA^{Phe} or human pre-tRNA^{Tyr} were incubated with cell extracts or TEV eluates, and formation of mature tRNA and/or tRNA exons was monitored by phosphorimaging.

Northern blot analysis. For tRNA detection, total RNA from tissues and cultured cells was isolated, subjected to gel electrophoresis and blotted on Hybond-N⁺ membranes. Blots were hybridized using [5' ³²P]-labelled DNA/LNA probes to detect Tyr-tRNA 5' and 3' exons.

Identification of Tyr-tRNA fragments. Total RNA from primary MEFs was separated by gel electrophoresis. The region containing RNA of 37–50 nucleotides in size was excised and sequenced using an Illumina platform. Reads were aligned to the mouse genome (<http://gtgnadb.ucsc.edu/Mmus/Mmuscul-by-locus-txt.html>) using bedtools (v. 2.16.2).

Received 6 March 2012; accepted 18 January 2013.

Published online 10 March 2013.

- Weitzer, S. & Martinez, J. The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. *Nature* **447**, 222–226 (2007).
- Ramirez, A., Shuman, S. & Schwer, B. Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme *in vivo*. *RNA* **14**, 1737–1745 (2008).
- Jain, R. & Shuman, S. Characterization of a thermostable archaeal polynucleotide kinase homologous to human Clp1. *RNA* **15**, 923–931 (2009).
- de Vries, H. et al. Human pre-mRNA cleavage factor II_m contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* **19**, 5895–5904 (2000).
- Minville-Sebastia, L., Preker, P. J., Wiederkehr, T., Strahm, Y. & Keller, W. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proc. Natl Acad. Sci. USA* **94**, 7897–7902 (1997).
- Holbein, S. et al. The P-loop domain of yeast Clp1 mediates interactions between CF IA and CPF factors in pre-mRNA 3' end formation. *PLoS ONE* **6**, e29139 (2011).
- Haddad, R. et al. An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. *Nucleic Acids Res.* **40**, 1226–1239 (2012).
- Ghazy, M. A. et al. The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic Acids Res.* **40**, 1214–1225 (2012).
- Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. *Cell* **117**, 311–321 (2004).
- Trotta, C. R. et al. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* **89**, 849–858 (1997).
- Zillmann, M., Gorovsky, M. A. & Phizicky, E. M. Conserved mechanism of tRNA splicing in eukaryotes. *Mol. Cell Biol.* **11**, 5410–5416 (1991).
- Zhao, C. et al. Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1B β . *Cell* **105**, 587–597 (2001).
- Sánchez-Carbente, M. R., Castro-Obregon, S., Covarrubias, L. & Narvaez, V. Motoneuronal death during spinal cord development is mediated by oxidative stress. *Cell Death Differ.* **12**, 279–291 (2005).
- Medana, I. M. & Esiri, M. M. Axonal damage: a key predictor of outcome in human CNS diseases. *Brain* **126**, 515–530 (2003).
- Atkin, J. D. et al. Properties of slow- and fast-twitch muscle fibres in a mouse model of amyotrophic lateral sclerosis. *Neuromuscul. Disord.* **15**, 377–388 (2005).

16. Yamasaki, S., Ivanov, P., Hu, G. F. & Anderson, P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* **185**, 35–42 (2009).
17. Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P. & Anderson, P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol. Cell* **43**, 613–623 (2011).
18. Son, E. Y. *et al.* Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* **9**, 205–218 (2011).
19. Lambert, P. F., Kashanchi, F., Radonovich, M. F., Shiekhata, R. & Brady, J. N. Phosphorylation of p53 serine 15 increases interaction with CBP. *J. Biol. Chem.* **273**, 33048–33053 (1998).
20. Dumaz, N. & Meek, D. W. Serine 15 phosphorylation stimulates p53 transactivation but does not directly influence interaction with HDM2. *EMBO J.* **18**, 7002–7010 (1999).
21. Chao, C. *et al.* Cell type- and promoter-specific roles of Ser18 phosphorylation in regulating p53 responses. *J. Biol. Chem.* **278**, 41028–41033 (2003).
22. Arber, S. *et al.* Requirement for the homeobox gene *Hb9* in the consolidation of motor neuron identity. *Neuron* **23**, 659–674 (1999).
23. Hanada, R. *et al.* Neuromedin U has a novel anorexigenic effect independent of the leptin signaling pathway. *Nature Med.* **10**, 1067–1073 (2004).
24. Tuck, A. C. & Tollervey, D. RNA in pieces. *Trends Genet.* **27**, 422–432 (2011).
25. Hurto, R. L. Unexpected functions of tRNA and tRNA processing enzymes. *Adv. Exp. Med. Biol.* **722**, 137–155 (2011).
26. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649 (2009).
27. Herbst, R., Iskratsch, T., Unger, E. & Bittner, R. E. Aberrant development of neuromuscular junctions in glycosylation-defective *Large^{myd}* mice. *Neuromuscul. Disord.* **19**, 366–378 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Meixner, M. Foong, T. Nakashima, H. C. Theussl, J. R. Wojciechowski, A. Bichl, the mouse pathology unit of the Universitätsklinikum

Hamburg-Eppendorf, and G. P. Resch for discussions and technical support. We also thank T. Buerckstuemmer for providing the pRV-NTAP vector. J.M.P. is supported by grants from the Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), the Austrian Ministry of Sciences, the Austrian Academy of Sciences, AustroMouse network of Genome Research in Austria (GEN-AU), Apoptosis systems biology applied to cancer and AIDS (ApoSys) and a European Research Council Advanced Grant from the European Union. J.M., S.W. and B.M. are supported by IMBA and GEN-AU (AustroMouse). T.H. is supported by the Japan Society for the Promotion of Science and the Astellas Foundation. J.K.I. was supported by National Institutes of Health (NIH) grant K99NS077435-01A1. M.G. is supported by grants from the German Research Foundation (DFG) (FG885 and GRK 1459), the Landesexzellenzinitiative Hamburg (Neurodapt). R.H. was supported by the Austrian Science Fund (P19223, P21667). C.J.W. is supported by the NIH (NS038253).

Author Contributions T.H. generated mutant mice, performed mouse phenotyping and developed cell lines with help from R.H. S.W. and B.M. performed all biochemical assays and northern blots. V.K. performed histological analysis. F.S., H.M., S.J.C. and A.Y. provided key reagents and technical help for generation of mutant mice. I.T. analysed tRNA fragment distributions. M.O. performed calorimetric experiments. B.J.W., J.I., K.C.E. and C.J.W. performed and helped with transdifferentiation experiments and patch clamping. R.H. performed immunostainings and assessment of NMJs and motor-neuron pathfinding. A.M. and A.Y. performed DRG explant cultures, and J.R. and R.K. carried out exon arrays. C.B. and M.G. performed histological and immunohistochemical staining of, and assessed, peripheral nerve and brain structures. J.M. and J.M.P. coordinated the project and wrote the manuscript.

Author Information Data have been deposited in the Gene Expression Omnibus under accession numbers GSE35924 and GSE39275. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. (javier.martinez@imba.oew.ac.at) or J.M.P. (josef.penninger@imba.oew.ac.at).

Structural visualization of key steps in human transcription initiation

Yuan He¹, Jie Fang², Dylan J. Taatjes³ & Eva Nogales^{1,2,4}

Eukaryotic transcription initiation requires the assembly of general transcription factors into a pre-initiation complex that ensures the accurate loading of RNA polymerase II (Pol II) at the transcription start site. The molecular mechanism and function of this assembly have remained elusive due to lack of structural information. Here we have used an *in vitro* reconstituted system to study the stepwise assembly of human TBP, TFIIA, TFIIB, Pol II, TFIIF, TFIIE and TFIIH onto promoter DNA using cryo-electron microscopy. Our structural analyses provide pseudo-atomic models at various stages of transcription initiation that illuminate critical molecular interactions, including how TFIIF engages Pol II and promoter DNA to stabilize both the closed pre-initiation complex and the open-promoter complex, and to regulate start-site selection. Comparison of open versus closed pre-initiation complexes, combined with the localization of the TFIIH helicases XPD and XPB, support a DNA translocation model of XPB and explain its essential role in promoter opening.

Accurate and regulated initiation of eukaryotic gene transcription represents a major step in gene regulation, requiring the coordinated activity of a large number of proteins and protein complexes. The basal transcriptional machinery includes RNA polymerase II (Pol II) along with a series of general transcription factors (GTFs) (TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH) that assemble into a ~2-megadalton (MDa) complex on core promoter DNA. This pre-initiation complex (PIC) is essential to direct accurate transcription start site (TSS) selection, promoter melting and Pol II promoter escape^{1–3}. Despite recent structural advances on Pol II^{4,5} and subcomplexes of the PIC⁶, the molecular assembly details of this essential complex remain elusive.

In vitro reconstitution of this process has provided a model for the sequential assembly pathway of transcription initiation. TFIID is the first factor specifically recruited to the promoter. This megadalton complex includes the TATA-binding protein (TBP), which is sufficient for basal transcription on TATA-box-containing promoters^{2,3,7}. TFIIA and TFIIB are then recruited, further stabilizing the interaction between TBP and promoter DNA. Next, Pol II, probably in association with TFIIF, adds to the growing PIC. Finally, TFIIE and TFIIH, which is required for DNA melting, are recruited to form the transcriptionally competent PIC^{2,3}.

Structural characterization of PIC assemblies is challenging and has been limited to a small number of electron microscopy (EM) studies^{8–10}. Crystallographic structures of individual components, combined with biochemical data, have led to a number of structural models for PIC subcomplexes, in either a closed or open-promoter conformation^{6,9,11,12}. In spite of this progress, important questions remain unanswered, such as how TFIIB and TFIIF serve complementary roles during the promoter opening process or how TFIIE positions TFIIH in a configuration capable of melting the DNA.

Here we present cryo-EM snapshots of PIC intermediates during sequential assembly (see Supplementary Video 1). A reconstitution system allowed us to localize each GTF within the cryo-EM structures, track the effect of each additional factor on the PIC, and ultimately reveal the network of protein–protein and protein–DNA interactions governing PIC assembly. Furthermore, by visualization of an open-promoter complex (OC) mimic, we have obtained new mechanistic

details concerning promoter melting. Together, our structures provide unprecedented insights into the molecular assembly, organization and functional roles of different GTFs during transcription initiation.

Stepwise assembly and visualization of human PIC

To structurally characterize the sequential assembly of GTFs necessary for human transcription initiation, we developed an *in vitro* system for reconstitution and purification of a simplified PIC, in which TBP substituted for TFIID, and that ultimately contained 31 polypeptides. Our promoter DNA contained TATA, BREu and BREd (upstream and downstream TFIIB recognition elements, respectively), and INR (initiator element) core promoter elements and was immobilized on streptavidin beads (Fig. 1a). After stepwise assembly of PIC intermediates by sequential incubation with the desired GTFs, stable complexes were released by restriction enzyme digestion. The effectiveness of this approach for structural characterization of the PIC intermediates was initially tested by single-particle EM of negatively

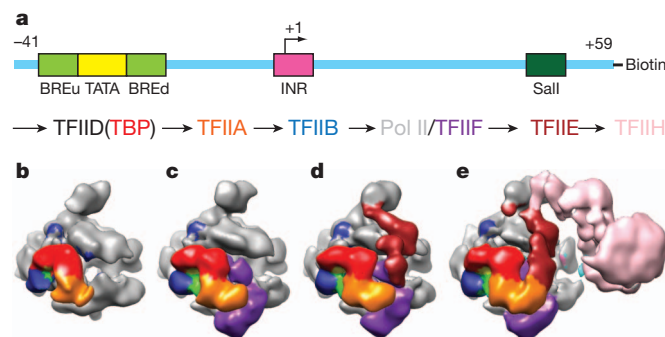


Figure 1 | Stepwise assembly of the human PIC. **a**, Reconstitution strategy for human PIC by sequential assembly. Schematic of the DNA highlighting the relative positions of the core promoter elements used and Sall restriction site (top). Colour scheme for the components of the PIC is shown at the bottom. **b–e**, Negative stain reconstructions of PIC assembly intermediates for TBP–TFIIA–TFIIB–DNA–Pol II (**b**), plus TFIIF (**c**), plus TFIIE (**d**), and plus TFIIH (**e**).

¹Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ²Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. ³Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80303, USA. ⁴QB3 Institute and Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

stained samples (Supplementary Fig. 1). This initial analysis allowed us to localize each GTF within the context of the full assembly (Fig. 1b–e), although it precluded the visualization of DNA. The stepwise purification approach enabled us to describe the effect of factor addition on the rest of the PIC, which cannot be achieved by studying individual factors or the complete PIC. The negative stain structures were then used as starting references to generate cryo-EM reconstructions of the PIC subcomplexes with improved resolution that allowed visualization of the DNA and accurate docking of existing crystal structures (Supplementary Figs 2–5).

Recruitment of Pol II onto promoter DNA

To start, we obtained the cryo-EM structure of a PIC subcomplex containing TBP, TFIIA, TFIIB, Pol II and core promoter DNA (Fig. 2a). Crystal structures of TBP–TFIIA–DNA¹³, TBP–TFIIB–DNA¹⁴ and yeast Pol II–TFIIB^{11,12,15} could be unambiguously docked into our density map as rigid bodies (Supplementary Fig. 2e). This procedure validated our cryo-EM structure while also allowing the localization of each factor to generate a pseudo-atomic model of the assembly. The visible portion of DNA accounts only for the upstream core promoter elements, which are stabilized by protein–DNA interactions, whereas the DNA downstream of the BRED lacks contact with the PIC and was not visualized due to its flexibility (Fig. 2a).

A yeast PIC model has been proposed previously, based on superimposing crystal structures using the common protein as an anchor point¹². Our data show that a simple pivoting of the carboxy-terminal cyclin fold domain of TFIIB around the amino-terminal one can explain the position of the TBP/TFIIA module in our map using the available crystal structures, without disrupting the interaction between the N-terminal cyclin fold of TFIIB and Pol II (Supplementary Fig. 7). This small discrepancy with the previous piece-wise model is unlikely to be due to differences between the human and yeast systems, but rather a re-organization with respect to individual crystal structures upon interaction of GTFs on the core promoter.

Effect of TFIIF on PIC assembly

According to the sequential assembly pathway, TFIIF is recruited to the promoter in association with Pol II². To understand its structural role during PIC assembly, we added TFIIF separately to our reconstituted system. By comparing the cryo-EM structures of PIC subcomplexes in the absence and presence of TFIIF, we identified additional protein densities appearing at two nearby locations, by the lobe and protrusion domains of Pol II (Fig. 2b). Importantly, the addition of TFIIF also resulted in the stabilization of the downstream DNA along the cleft of Pol II, in a position that is distinct from a previously proposed model¹² (Supplementary Fig. 8). Thus, TFIIF is required for the engagement of DNA by Pol II within the context of a closed PIC.

The crystal structure of the human TFIIF dimerization domain¹⁶ could be unambiguously fitted into the new density ascribed to TFIIF by the lobe domain of Pol II using rigid-body docking, only slightly shifted from previous models that were based on crosslinking data^{17,18} (Fig. 2c and Supplementary Fig. 3e). No obvious density was observed for the arm domain of RAP74, which extends about 45 Å from the end of the RAP74 barrel¹⁶, indicating that it is mobile at this stage of PIC assembly. A small clash between the RPB2 lobe and the RAP74 α 1 helix can be explained by a reorganization of this element in the context of the PIC (see Supplementary Fig. 9 for details).

In addition to interacting with Pol II, our cryo-EM structure indicates that a region of TFIIF contacts the BRED directly (Fig. 2c, d). We propose that this region corresponds to the C-terminal winged-helix (WH) domain of RAP30, based upon the following: (1) the size of this additional TFIIF density is consistent with the RAP30 WH domain; (2) RAP30 has been shown to crosslink to BRED¹⁹ and its C-terminal WH domain has been identified to be in direct contact with the protrusion domain of Pol II¹⁸; (3) RAP30, rather than RAP74, is

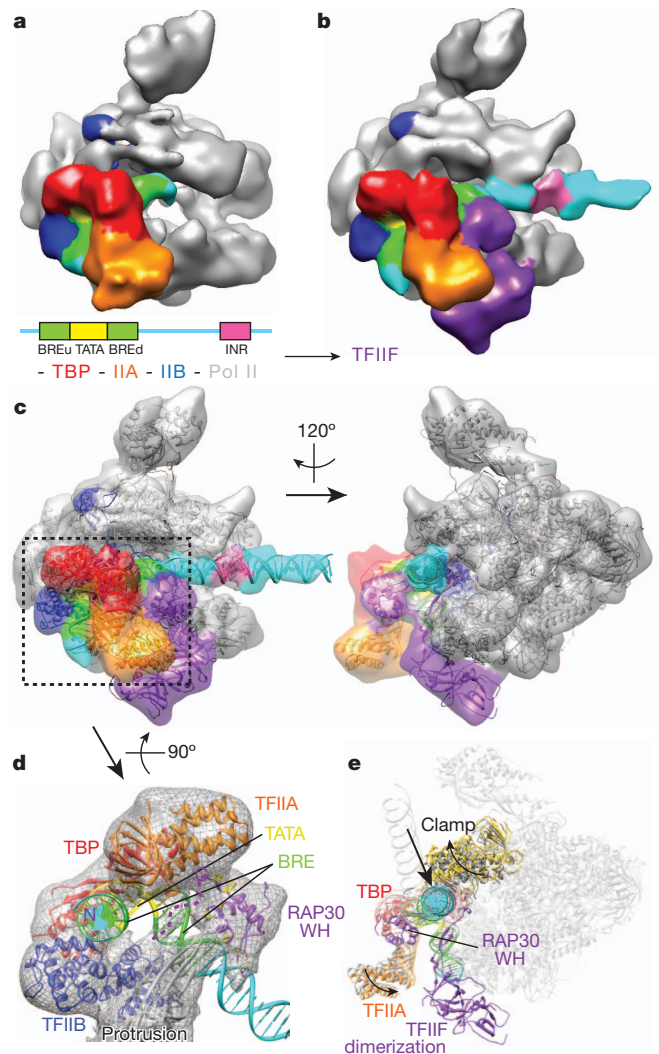


Figure 2 | TFIIF engagement triggers a concerted conformational change in the PIC. **a, b**, The positions of TBP, TFIIA, TFIIB and DNA promoter elements relative to Pol II are indicated in the cryo-EM reconstruction in the absence (**a**) or presence of TFIIF (**b**). TFIIF stabilizes promoter DNA (cyan). **c**, Crystal structures for TBP–TFIIA–DNA (Protein Data Bank (PDB) accession 1NVP), TBP–TFIIB–DNA (PDB 1C9B), Pol II–TFIIB (PDB 4BBR), RAP30/74 dimerization domain (PDB 1F3U¹⁶), RAP30 WH domain (PDB 1BBY⁴⁹), and modelled B-form DNA (globally bent between -23 and $+7$ by 18°) are shown docked into transparent EM densities. The mobile clamp of Pol II is docked as a separate domain. **d**, Bottom view showing the presence of a nucleoprotein complex by the upstream core promoter elements. The DNA densities have been segmented out for clarity. Perturbed residues within the RAP30 WH domain during DNA titrations are coloured in gold⁴⁹. The possible path for the RAP30 N terminus is highlighted with dotted purple lines. DNA is shown in ribbon representation. **e**, Position of moving structural elements before (grey) and after (coloured) TFIIF binding. The rest of the PIC components are shown in transparency and major structural rearrangements are depicted by arrows.

required for accurate transcription initiation²⁰ and deletion mutants of RAP30's WH domain are lethal in yeast^{18,21}. This WH domain therefore contributes to a unique nucleoprotein complex formed by TBP, TFIIA and the TFIIB cyclin fold, as they contact the core promoter elements upstream of the INR, which is further stabilized by the protrusion domain of Pol II (Fig. 2d and Supplementary Fig. 10).

Comparison of structures shown in Fig. 2a and b strongly suggests that the overall effect of TFIIF on the assembling PIC is a clear stabilization of the DNA along the Pol II cleft. Given its location, we propose that the RAP30 WH domain has an essential role in positioning the flexible promoter DNA downstream of BRED along the Pol II cleft, thus facilitating subsequent steps in the promoter melting

process. However, a contribution to this DNA stabilization by the dimerization domain is also possible. Correct positioning of the DNA by TFIIF is consistent with its role in promoter opening and TSS selection^{22–24}. Our structures also revealed the opening of the Pol II clamp domain as it accommodates the downstream DNA (Fig. 2e and Supplementary Fig. 11). In addition, we observe further rotation of the TBP–TFIIA–TFIIB subcomplex with respect to previously proposed models, positioning it even closer to Pol II (Fig. 2e and Supplementary Fig. 7).

TFIIE-containing PIC and DNA contacts

The addition of TFIIE to the growing PIC resulted in new protein density that connects TFIIF with the stalk domain of Pol II (Fig. 3a). The resolution of this reconstruction (11 Å) was the highest obtained for any of the complexes studied, indicating that TFIIE stabilizes the PIC. The density corresponding to TFIIE, however, was the least well-defined element according to local resolution calculations (Supplementary Fig. 4f), which may be due to flexible connections between the WH domains predicted within the TFIIE structure⁶ (Fig. 3a, b). One end of TFIIE associates with the stalk of Pol II by interacting with the RPB7 L45 loop (Fig. 3c), which has been predicted to stabilize the OC and the deletion of which completely abolished transcription²⁵. Also consistent with the positioning of TFIIE in our PIC structure, a zinc ribbon domain within the archaeal homologue of TFIIE was found to be located near the base of the stalk domain of the polymerase²⁶. Away from the stalk, the TFIIE density contacts the Pol II clamp domain to interact ultimately with the WH domain of TFIIF. A model of the three WH domains within TFIIE interacting with elements of the clamp head has been proposed based on

crosslinking studies⁶ (Fig. 3c). Although the model cannot fit the EM density perfectly, the overall path of the three tandem WH domains in the model follows the elongated TFIIE cryo-EM density and ends by directly contacting the RAP30 WH domain (Fig. 3b). Therefore, a continuous chain of four WH domains appears to link the Pol II clamp region with the TBP–TFIIA–TFIIB–DNA subcomplex, preventing DNA from leaving the cleft.

Our 11 Å resolution reconstruction of the PIC containing TFIIE starts to reveal the major and minor grooves of the promoter DNA (Fig. 3d), allowing us to model its path. We found that linear B-form DNA could not be accommodated into the DNA density (Supplementary Fig. 12), requiring instead a smooth bend of 18° between positions –23 and +7 that fitted both the path and groove features of the EM density. Interestingly, a hypersensitivity region around –6 position²⁷ locates at one of the downstream DNA–Pol II interfaces as discussed below.

We observed two protein contacts with the downstream DNA. One connection involves the 3-strand β -sheet below the clamp head whereas the other is mediated by a 2-helix bundle at the tip of the RPB5 jaw (Fig. 3d). Interestingly, these are the only two positively charged protein surfaces on Pol II along the path of the downstream DNA (Supplementary Fig. 12). The INR element is sandwiched precisely between these two protein–DNA contacts, an arrangement that may be relevant in promoter melting at the correct position in the DNA. The slightly open clamp conformation seen upon DNA placement onto the cleft after TFIIF addition is probably due to the interaction of the DNA with the clamp head β -sheet (Fig. 3d and Supplementary Fig. 11b, c).

The spacing between the TATA box and the TSS can vary between species, whereas the region within promoter DNA that is melted during transcription initiation is ~20 base pairs (bp) downstream of TATA²⁸. We inferred the approximate position of flexible elements within TFIIB and TFIIF by docking their crystal structures as rigid bodies within our cryo-EM density. Importantly, we find that both the TFIIB linker helix and the TFIIF arm domain align with the promoter melting start site (Supplementary Fig. 13). This arrangement is consistent both with the proposed role of the linker helix of TFIIB in promoter opening¹² and with the crosslinking of the arm domain of RAP74 to the TFIIB linker near the active site^{29,30}, as well as with the suppression of the TSS defect of TFIIB mutations by a mutant within the arm domain of TFIIF³¹. In our rigid-body fitting, the linker helix of TFIIB overlaps with the DNA in our model, suggesting a rearrangement of the helix relative to the clamp domain at this stage in the PIC assembly. Finally, the tip of the TFIIF arm domain contains seven positively charged residues, whereas four positively charged residues are present on the side of the TFIIB linker helix that faces the DNA (Supplementary Fig. 13). The juxtaposition of these domains within the melting start site is consistent with their direct role in DNA interactions.

The structural features of our Pol II-based PIC model are probably conserved with Pol I and Pol III, the two other RNA polymerases in eukaryotes. A side-by-side comparison of our Pol II-based PIC model with a cryo-EM structure of native Pol III agrees with this hypothesis³² (Supplementary Fig. 14).

Transition to the open-promoter complex

It is well established that the PIC remains stably associated during transcription initiation until Pol II undergoes promoter escape². Preceding this step, however, Pol II needs to transition into an OC in which the melted single-stranded DNA is inserted into the active site. To gain structural insight into the transition from a closed to an open promoter complex, we generated a ‘functional mimic’ of the PIC in its open conformation by modifying the promoter substrate used to form the closed PIC (Fig. 4a). We replaced the segment of DNA containing the INR element with a 3′-tailed sequence previously used to create an arrested transcription state in yeast Pol II³³. We matched

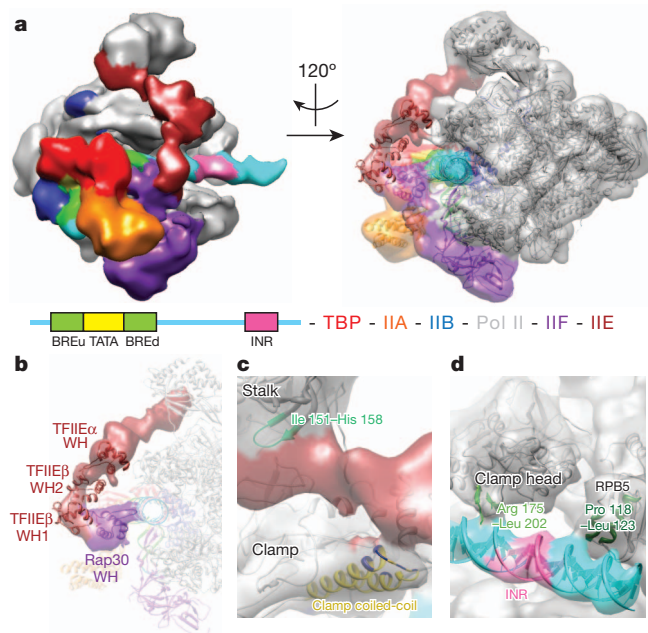


Figure 3 | Stabilization of the PIC in the closed conformation by TFIIE.

a, Segmentation of the cryo-EM reconstruction of human PIC containing TFIIE (left), and docking of existing crystal structures (right). **b**, Same view as the right panel in **a** depicting the cryo-EM density corresponding to TFIIE and the RAP30 WH domain. A chain of four WH domains formed by the C terminus of RAP30 and both subunits of TFIIE (modelled based on crosslinking data⁶) can be roughly fitted into the cryo-EM density. **c**, Regions of contact between TFIIE and Pol II. The clamp and stalk domain of Pol II are shown in grey, TFIIE in maroon. The rest of the PIC components are shown in transparency. The clamp coiled-coil domain is shown in gold and the residues that crosslink to TFIIE⁵⁰ are coloured in navy blue. **d**, EM density corresponding to promoter DNA is shown together with the PIC ribbon model. Structural elements making direct contacts with the promoter DNA are depicted.

the arrested position of Pol II on the template exactly to the TSS used in our studies, thereby creating a Pol II–nucleic acid complex containing only ~5 nucleotides at the active site, while still containing upstream core promoter elements available for assembling the rest of the PIC. We found that TFIIE had a higher affinity for the OC, as excess TFIIE had to be used to saturate the closed PIC, but not the open state mimic. Interestingly, excess TFIIE was no longer required in the context of the closed PIC when TFIIH was also included (see below), in agreement with previous studies suggesting cooperative binding of TFIIE and TFIIH within the PIC^{2,34,35}.

The reconstruction of the OC mimic resembled that of the PIC in the closed conformation, with all the GTFs remaining at identical positions (Fig. 4b and Supplementary Figs 4e, 5e and 10b–d). This

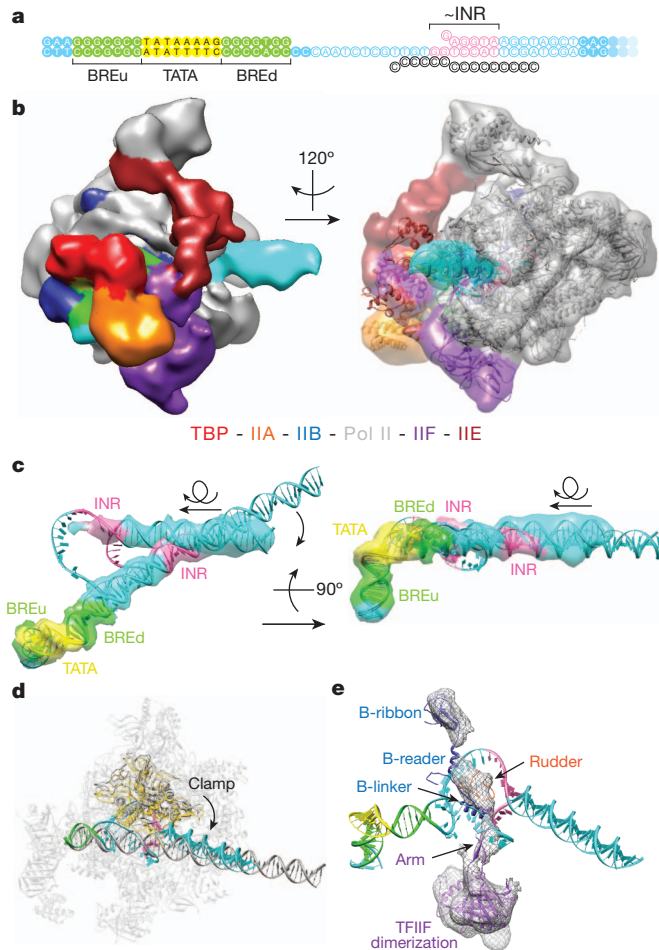


Figure 4 | Conformational rearrangements of the PIC upon promoter opening. **a**, Nucleic acid scaffold used to generate a mimic of the OC. Filled and open circles correspond, respectively, to the core promoter used in Fig. 1a, and to a replacement sequence containing a 3'-tailed sequence previously designed to generate an arrested Pol II³³. The schematic indicates, in pink, the position that would correspond to the INR. **b**, Segmentation of the cryo-EM reconstruction of the human PIC in the open state (left), and docking of existing crystal structures, together with a modelled DNA bubble¹² (right). GTFs adopt the same architecture as in the closed PIC. **c**, EM densities of promoter DNA in the closed PIC and OC were segmented and superimposed with respect to Pol II. Movements of the DNA between the two states are shown by arrows. The rotation accompanying translocation of the downstream DNA occurs within the plane of the view shown on the left panel (and thus perpendicular to view shown on the right). **d**, Structures before (grey) and after (coloured) promoter opening are shown using fitted crystal structures. The Pol II clamp comes down over the open bubble, in a conformation similar to that seen for the elongation state³⁷. **e**, Segmented cryo-EM density showing the now visible arm domain of TFIIF connecting the rest of TFIIF with the rudder of RPB1. Crystal structures of TFIIB and TFIIF were docked as rigid bodies into the cryo-EM density.

finding is consistent with the prevalent hypothesis that the PIC assembled at the promoter remains intact until promoter escape². In contrast, the downstream DNA adopted a conformation previously observed for elongating Pol II, indicating that the template strand was inserted through the positively charged cleft into the active site⁴ (Fig. 4b, c). The single-stranded segments are invisible at our resolution or not present (non-template strand). As a reasonable model, the bubble depicted in Fig. 4 has been derived from a previous model based on FRET studies on the yeast system³⁶.

When the position and orientation of the downstream DNA is compared between the closed PIC and the OC mimic, it is clear that there is a change in orientation concomitant with the insertion of the downstream DNA into the active site (Fig. 4c), indicating that the DNA rotates on a plane as it translates, while maintaining a point of contact between the DNA and RPB5 that corresponds to one of the two contacts present in the closed state (the one downstream of the INR).

Other than the repositioning of the DNA within the active site, two main differences were observed upon comparison of the OC mimic and the closed PIC structures. First, the clamp domain in the open state moves down to engage the open DNA bubble, adopting the conformation observed in the elongation state³⁷ (Fig. 4d and Supplementary Fig. 11d). Thus, the clamp domain completes an open to closed transition throughout the process of PIC assembly and promoter opening (Supplementary Fig. 11), a cycle also reported for the bacterial system³⁸. Second, an additional protein density now extends from the bottom of the clamp and connects to the dimerization domain of TFIIF (Fig. 4e). Rigid-body fitting of crystal structures suggests that this density corresponds to the stabilized rudder of Pol II and the arm domain of TFIIF. We propose that these elements interact with each other as the clamp closes down over the melted DNA. Interestingly, this proposed interaction would prevent re-annealing of the melted DNA. The TFIIB linker helix is near this position and probably participates in the promoter melting process as well. This proposal is consistent with our hypothesis that the flexible TFIIB linker helix and the TFIIF arm domain act together in promoter opening (Supplementary Fig. 13). Thus, our structure and pseudo-atomic model provide a possible explanation for the enigmatic role of TFIIF in promoter opening and TSS selection^{22,23}.

Positioning of XPB for DNA translocation

To gain insight into the natural promoter opening process carried out by TFIIH, we used the same purification strategy used for the previous closed PIC subcomplexes but included the purified, endogenous 10-subunit human TFIIH complex as a last step before elution. Given the scarcity of purified human TFIIH, this study was limited to negative stained samples, which require less material. The three-dimensional reconstruction of the TFIIH-containing PIC showed a substantial additional density extending away from Pol II, consistent with the large molecular mass of TFIIH (0.45 MDa) (Figs 1e and 5a and Supplementary Fig. 6e). Surprisingly, only two contacts are observed between TFIIH and the rest of the PIC. One is with the Pol II's stalk domain, at the site of interaction with TFIIE. The other contact probably involves the interaction of TFIIH directly with the downstream DNA. Although the DNA is not visible in this negative stain reconstruction, its position can be extrapolated from the cryo-EM structure of the PIC containing TFIIE (Fig. 5a).

The CAK subcomplex (CDK7–cyclin H–MAT1) of TFIIH, which phosphorylates the C-terminal domain (CTD) of RPB1, is missing from our PIC reconstruction based on comparison with a recent EM study of yeast TFIIH³⁹. When we analysed images of free human TFIIH, an additional density that could accommodate the mass of the CAK subcomplex appeared highly mobile, in agreement with the yeast TFIIH data³⁹ (Supplementary Figs 15 and 16). Interestingly, when this new density, which fits the crystal structures of CDK7 and cyclin H, is placed in the context of the full PIC, it faces towards the CTD of RPB1 (Supplementary Fig. 16).

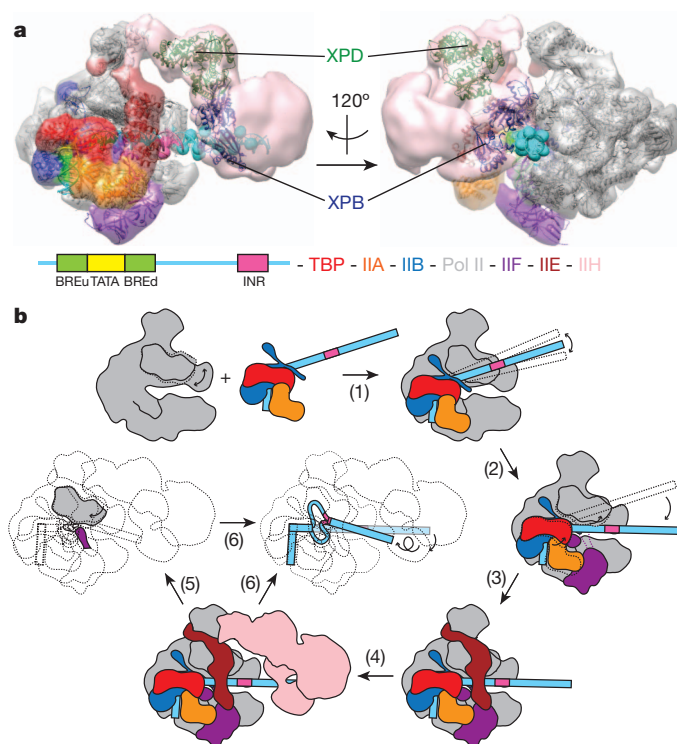


Figure 5 | Positioning of TFIIF helicases and model of PIC assembly and promoter opening. **a**, Negative-stain structure of the full human PIC in the closed conformation (as in Fig. 1e). EM density corresponding to the TFIIF core complex, lacking any visible CAK subcomplex, is coloured in pink. A homology model for XPB⁶ (navy blue) and the crystal structure of XPD (PDB 3CRV, dark green) are shown docked into the core TFIIF density. The docking suggests different roles for the XPB and XPD helicases in promoter opening. DNA phosphates crosslinked to XPB are indicated by pink or cyan spheres⁴¹. **b**, Schematic of PIC assembly and promoter melting. Pol II is recruited through interaction with TFIIB to the promoter, which is engaged by TBP-TFIIA-TFIIB (1). TFIIF stabilizes the TBP-TFIIA-TFIIB-Pol II protrusion interaction hub and also positions the downstream DNA onto the cleft, forcing the clamp to swing into a slightly open state (2). TFIIE binding further stabilizes the PIC by interacting with the Pol II stalk, the clamp, and with TFIIF on the other side of the Pol II cleft (3). The TFIIE-containing PIC serves as the platform for TFIIF binding and correctly positions XPB downstream of the INR element (4). During strand separation, the clamp domain starts to swing down. The arm domain of TFIIF comes close to the Pol II rudder and the TFIIB B-linker. Stabilization of these interactions forms a physical barrier for DNA re-annealing (5). During promoter opening, the translocase activity of XPB would 'screw in' the DNA towards the Pol II active site, leading to a Pol II open-promoter state ready for RNA synthesis (6).

The reconstruction of the TFIIF-containing PIC allowed us to dock the crystal structure of XPD⁴⁰ and a homology model of XPB⁶ (Fig. 5a and Supplementary Fig. 6e). XPD is positioned in close proximity to TFIIE and the Pol II stalk, but away from DNA, consistent with a scaffolding role in transcription initiation⁷. On the other hand, XPB docked directly on the downstream DNA path, between the +10 and +20-bp position relative to the TSS (Fig. 5a). This position is consistent with previous crosslinking data using purified TFIIF⁴¹, but inconsistent with a recent crosslinking study using overexpressed XPB in extracts, in which XPB was proposed to be positioned closer to the TFIIE WH domains and the INR element⁶. This result might reflect an alternative position of this protein during the assembly of the PIC, a distinct position of XPB on the DNA when out of the TFIIF complex, or the effect of other factors like TFIID and mediator on PIC organization. The position of XPB that we observed within the TFIIF density, together with the movement of the downstream DNA inferred from comparison of our reconstructions of the closed and open states of Pol II, suggests how XPB could act as a DNA translocase. A translocase

model for XPB has previously been proposed⁶, but our structure now shows XPB positioned further downstream, leaving enough space around the INR element for it to be melted during this process.

We believe that the position of XPB suggests a DNA insertion process in which, as XPB walks on the DNA away from the rest of PIC, the DNA would be translocated in the opposite direction and pushed into the Pol II cleft while maintaining a point of contact with RPB5 (which starts involving the DNA just downstream of the INR in the closed complex). This happens concomitantly with a rotation of the DNA, with the RPB5 contact probably serving as a pivot point. As XPB walks on the DNA helix, it would generate supertwist that would be relaxed by unwinding. Although this unwinding cannot happen in the DNA that is tightly wrapped and stabilized by the TBP-TFIIA-TFIIB-RAP30-protrusion module, it would be facilitated and/or stabilized in the DNA region between the BREd and the INR, where the arm domain of RAP74 makes contact with the Pol II rudder and regions of TFIIB.

Concluding remarks

The combination of structures described here provides unprecedented mechanistic insight into the stepwise assembly of the human PIC, defining key protein-protein and protein-DNA interactions important for PIC function (Fig. 5b and Supplementary Video 1). Our structures reveal the location and role of RAP30 WH domain within an essential upstream nucleoprotein subcomplex. Its critical function in structurally stabilizing the whole PIC is highlighted by our direct visualization of the DNA as it is positioned along the Pol II cleft upon TFIIF binding. We also show a direct interaction between the arm module of the TFIIF dimerization domain and the rudder domain of Pol II upon formation of the OC, leading to a direct mechanistic model of how this TFIIF element facilitates and/or maintains strand separation concomitant with the closing down of the clamp domain of Pol II. Our structures show how two essential factors, TFIIB and TFIIF, come together at critical locations for their activity in the context of a full PIC. Our studies also reveal how TFIIF, because of its large size, can simultaneously interact with TFIIE at the base of the Pol II stalk and position XPB on downstream DNA.

Our studies of the closed PIC and an OC mimic illuminate the structural transitions necessary during the process of promoter melting. The apparent movement of downstream DNA observed when comparing the closed PIC and OC structures, together with the positioning of XPB on the downstream DNA, suggests how XPB could act as a DNA translocase to thread approximately 10 bp of downstream double-stranded DNA into the cleft. This translocating activity would push against the stably bound upstream DNA around the TATA box to induce negative supercoiling near the TSS. We find that the TFIIB linker helix and the TFIIF arm domain align with each other at the promoter melting start site, probably to facilitate the separation of the two strands. Once promoter DNA melting is further extended and the Pol II clamp closes down, the TFIIB linker helix and the TFIIF arm domain work together with the Pol II rudder to maintain the upstream edge of the DNA bubble.

Finally, the arrangement of components within our PIC structure is compatible with existing structural models that include the large, multi-subunit mediator and TFIID complexes¹⁰. Future structural studies with mediator and/or TFIID will yield further insight regarding the regulation of PIC assembly and function. In summary, this work provides the structural framework needed to integrate biochemical and structural data into a unified mechanistic understanding of transcription initiation.

METHODS SUMMARY

Protein expression and purification. TBP, TFIIA, TFIIB, TFIIE and TFIIF were recombinantly expressed and purified from *Escherichia coli*. Pol II and TFIIF were immunopurified from HeLa cell nuclear extracts⁴². The design of the DNA construct was based on the SCP⁴³, with a SalI restriction enzyme site introduced

downstream of the INR element. PIC complexes were assembled according to an *in vitro* transcription protocol⁴² with minor modifications (see Methods). The reactions were incubated with magnetic streptavidin T1 beads (Invitrogen) and the desired complexes were eluted by SalI digestion.

EM and image analysis. Data collection and image processing were conducted using the Leginon data collection software⁴⁴ and the Appion electron microscopy processing environment⁴⁵, respectively. Three-dimensional maps were calculated using libraries from the EMAN2 and SPARX software packages^{46,47}. Volume segmentation, automatic rigid-body docking, figure and movie generation were performed using UCSF Chimera⁴⁸.

Full Methods and any associated references are available in the online version of the paper.

Received 26 September 2012; accepted 7 February 2013.

Published online 27 February 2013.

- Matsui, T., Segall, J., Weil, P. A. & Roeder, R. G. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* **255**, 11992–11996 (1980).
- Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
- Goodrich, J. A., Cutler, G. & Tjian, R. Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* **84**, 825–830 (1996).
- Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proc. Natl Acad. Sci. USA* **104**, 12955–12961 (2007).
- Cramer, P. *et al.* Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).
- Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Struct. Mol. Biol.* **19**, 788–796 (2012).
- Thomas, M. C. & Chiang, C. M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178 (2006).
- Andel, F. III, Ladurner, A. G., Inouye, C., Tjian, R. & Nogales, E. Three-dimensional structure of the human TFIID-IIA-III complex. *Science* **286**, 2153–2156 (1999).
- Chung, W. H. *et al.* RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol. Cell* **12**, 1003–1013 (2003).
- Bernecky, C., Grob, P., Ebmeier, C. C., Nogales, E. & Taatjes, D. J. Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. *PLoS Biol.* **9**, e1000603 (2011).
- Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010).
- Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
- Bleichenbacher, M., Tan, S. & Richmond, T. J. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *J. Mol. Biol.* **332**, 783–793 (2003).
- Tsai, F. T. & Sigler, P. B. Structural basis of preinitiation complex assembly on human pol II promoters. *EMBO J.* **19**, 25–36 (2000).
- Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437–440 (2013).
- Gaiser, F., Tan, S. & Richmond, T. J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J. Mol. Biol.* **302**, 1119–1127 (2000).
- Chen, Z. A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
- Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
- Robert, F., Forget, D., Li, J., Greenblatt, J. & Coulombe, B. Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. *J. Biol. Chem.* **271**, 8517–8520 (1996).
- Tyree, C. M. *et al.* Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes Dev.* **7**, 1254–1265 (1993).
- Tan, S., Garrett, K. P., Conaway, R. C. & Conaway, J. W. Cryptic DNA-binding domain in the C terminus of RNA polymerase II general transcription factor RAP30. *Proc. Natl Acad. Sci. USA* **91**, 9808–9812 (1994).
- Ghazy, M. A., Brodie, S. A., Ammerman, M. L., Ziegler, L. M. & Ponticelli, A. S. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol. Cell. Biol.* **24**, 10975–10985 (2004).
- Yan, Q., Moreland, R. J., Conaway, J. W. & Conaway, R. C. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* **274**, 35668–35675 (1999).
- Forget, D. *et al.* RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. *Proc. Natl Acad. Sci. USA* **94**, 7150–7155 (1997).
- Orlicky, S. M., Tran, P. T., Sayre, M. H. & Edwards, A. M. Dissociable Rpb4-Rpb7 subassembly of *ma* polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem.* **276**, 10097–10102 (2001).
- Grohmann, D. *et al.* The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell* **43**, 263–274 (2011).
- Burawski, S., Söpta, M., Greenblatt, J. & Sharp, P. A. RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. *Proc. Natl Acad. Sci. USA* **88**, 7509–7513 (1991).
- Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
- Chen, H. T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**, 169–180 (2004).
- Freire-Picos, M. A., Krishnamurthy, S., Sun, Z. W. & Hampsey, M. Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. *Nucleic Acids Res.* **33**, 5045–5052 (2005).
- Sun, Z. W. & Hampsey, M. Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **92**, 3127–3131 (1995).
- Fernández-Tornero, C. *et al.* Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.* **29**, 3762–3772 (2010).
- Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
- Goodrich, J. A. & Tjian, R. Transcription factors IIE and IIIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77**, 145–156 (1994).
- Conaway, R. C. & Conaway, J. W. General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190 (1993).
- Andrecka, J. *et al.* Nano positioning system reveals the course of upstream and non-template DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res.* **37**, 5803–5809 (2009).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Chakraborty, A. *et al.* Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**, 591–595 (2012).
- Gibbons, B. J. *et al.* Subunit architecture of general transcription factor TFIIF. *Proc. Natl Acad. Sci. USA* **109**, 1949–1954 (2012).
- Fan, L. *et al.* XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**, 789–800 (2008).
- Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1421 (2000).
- Revyakin, A. *et al.* Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev.* **26**, 1691–1702 (2012).
- Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nature Methods* **3**, 917–922 (2006).
- Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
- Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
- Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
- Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
- Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
- Groft, C. M., Uljon, S. N., Wang, R. & Werner, M. H. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc. Natl Acad. Sci. USA* **95**, 9117–9122 (1998).
- Chen, H. T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature Struct. Mol. Biol.* **14**, 696–703 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Inouye for providing us with recombinant TFIIF and TFIIE; P. Grob and T. Houweling for electron microscopy and computer support, respectively; T. Goddard for help with Chimera; and members of the Nogales laboratory for technical advice on image processing. We are thankful to J. Kadonaga, J. Goodrich and M. Cianfrocco for their comments on the manuscript. We thank P. Cooper both for biochemical advice and for her comments on the manuscript. This work was funded by NIGMS (GM63072 to E.N.) and by NCI (CA127364 to D.J.T.). E.N. is a Howard Hughes Medical Institute Investigator.

Author Contributions Y.H. designed and carried out the experiments; J.F. and D.J.T. provided essential reagents; Y.H. and E.N. analysed the data and wrote the paper.

Author Information Cryo-EM density maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-2304 (TBP-TFIIA-TFIIB-DNA-Pol II), EMD-2305 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF), EMD-2306 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE) and EMD-2307 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE in the OC mimic state). Negative stain EM density maps have been assigned accession numbers EMD-2308 (TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF-TFIIE-TFIIH) and EMD-2309 (apo TFIIF). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.N. (ENogales@lbl.gov).

METHODS

PIC assembly and purification. TBP, TFIIA, TFIIB, TFIIE and TFIIF were recombinantly expressed and purified from *Escherichia coli*. Pol II, TFIID and TFIIF were immunopurified from HeLa cell nuclear extracts following previously established protocols^{12,51}. The design of the DNA construct was based on the SCP⁴³, except that a BREu element was introduced upstream of the TATA box¹⁴ and a SalI restriction enzyme site was included downstream of the INR element for purification purposes (template1, 5'-ACTGGGAAGTCGACCGG TCCGTAGGCACGTCTGCTCGGCTCGAGTGTTCGATCGCGACTGAGGAC GAACGCGCCCCACCCCTTTTATAGCGCCCTTC-3'; nontemplate1, 5'-GAAGGGCGCCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGC GATCGAACACTCGAGCCGAGCAGACGTGCCTACGGACCGGTGCGACTT CCCAGT-3'). The nucleic acid scaffold that was used to generate the PIC in the open conformation was designed by modification of the promoter substrate used to form the closed PIC. An RNA–DNA duplex beyond 7 bp has been proposed to be the trigger for TFIIB release and promoter escape^{12,52}. Thus, we replaced the segment of DNA containing the INR element with a 3'-tailed sequence previously used to create an arrested transcription state in yeast Pol II³³. We matched the arrested position of Pol II on the template exactly to the TSS used in our studies, thereby creating a Pol II–nucleic acid complex containing only about 5 nucleotides at the active site while still containing upstream core promoter elements available for assembling the rest of the PIC (template2, 5'-ACTGGGAAGTCGACCGGT CCGTAGGCACGTCTGCTCGGCTCGAGTGTTCGATCGCGACTGAGGAC GAACGCGCCCCACCCCTTTTATAGCGCCCTTC-3'; nontemplate2, 5'-G AAGGGCGCCTATAAAAGGGGGTGGGGGTT-3'; nontemplate3, 5'-GAGGT AAGCTAGCTCACTCGAGCCGAGCAGACGTGCCTACGGACCGGTGCGACT TCCAGT-3'). A biotin tag was engineered at the 5' end of both template strands (Integrated DNA Technologies). The duplexed DNA was generated by annealing the template strand with equimolar amounts of single-stranded non-template DNA at a final concentration of 50 μ M in water. The annealing reaction was carried out at 100 °C for 5 min and gradually cooled down to room temperature within 1 h.

PICs in the closed conformation were assembled according to an *in vitro* transcription protocol⁴² with minor modifications. The assembly buffer contained 12 mM HEPES, pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM MgCl₂, 60 mM KCl, 1 mM DTT, 0.05% NP-40, 2.5 ng μ L⁻¹ dI-dC, 10 μ M ZnCl₂. The following purified proteins and nucleic acids were sequentially added into the assembly buffer: Pol II, TFIIB, TBP/TFIIA, DNA (template1–nontemplate1), TFIIF and TFIIE at final concentrations of 185 nM, 3.6 μ M, 370 nM, 50 nM, 289 nM and 370 nM, respectively. The assembly reaction was kept at 37 °C for an additional 5 min whenever a new factor was added. The reaction was incubated at 28 °C for 15 min using a 1:10 dilution of the magnetic streptavidin T1 beads (Invitrogen) which had been equilibrated with the assembly buffer. Following washing of the beads three times using a washing buffer (10 mM HEPES, 10 mM TRIS, pH 7.9, 5% glycerol, 5 mM MgCl₂, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 5 μ M ZnCl₂), TFIIF at a final concentration of 100 nM was incubated with the beads in assembly buffer at 37 °C for an additional 5 min. Following a single additional wash of the beads using washing buffer, the desired complex was eluted by incubating the beads at 28 °C for 1 h with digestion buffer containing 10 mM HEPES, pH 7.9, 5% glycerol, 10 mM MgCl₂, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 1 unit μ L⁻¹ BSA-free SalI-HF (New England Biolabs). The various PIC intermediates were generated by including just the factors of interest during the assembly process described above. For preparing TBP–TFIIA–TFIIB–DNA–Pol II–TFIIF–TFIIE complex, extra TFIIE was added afterwards to the purified PIC at a final concentration of 100 nM.

It was not possible for us to reconstitute an open complex using either a mismatch DNA bubble, probably due to failure of efficiently and specifically positioning Pol II on the bubble, or a nucleic acid scaffold containing an RNA primer, probably because an RNA–DNA duplex of over 7 bp in length has been proposed to be the trigger for TFIIB release and promoter escape^{12,52}. PIC in the open conformation was assembled similarly, except for the following changes. An arrested Pol II on the open promoter nucleic acid scaffold was first prepared by incubation at 28 °C for 1 h of Pol II and DNA (template2–nontemplate3) at final concentrations of 300 nM and 80 nM, respectively, in the arresting buffer containing 12 mM HEPES, pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM MgCl₂, 60 mM KCl, 1 mM DTT, 0.05% NP-40, 2.5 ng μ L⁻¹ dI-dC, 1:100 dilution of RNasin Ribonuclease inhibitor (Promega), and 2 mM CTP. The following purified proteins and nucleic acid were sequentially added into the arrested Pol II reaction above: nontemplate2, TBP/TFIIA, TFIIB, TFIIF and TFIIE at final concentrations of 200 nM, 370 nM, 3.6 μ M, 289 nM and 370 nM, respectively. The desired open-promoter complex was then purified in the same manner as the closed complexes above.

Purified PIC complexes were crosslinked after elution by incubation with glutaraldehyde at a final concentration of 0.05%, on ice and under very low illumination conditions, for 5 min, then immediately used for EM sample preparation (either negative stain or cryo-plunging).

Electron microscopy. Negative stain samples of PIC complex and of free TFIIF were prepared using 400 mesh copper grid containing a continuous carbon supporting layer. The grid was plasma cleaned for 10 s immediately before sample deposition using a Solarus plasma cleaner (Gatan) equipped with 75% argon/25% oxygen. An aliquot (3 μ L) of the purified sample (~50 nM) was placed onto the grid and allowed to absorb for 5 min at 100% humidity in a homemade humidity chamber kept under very low illumination conditions. It was subsequently stained by five successive 75 μ L drops of 2% (w/v) uranyl formate solution, rocking 10 s on each drop followed by blotting till dryness. Data collection was performed using a Tecnai F20 Twin transmission electron microscope operating at 120 keV at a nominal magnification of $\times 80,000$ (1.37 Å per pixel). The data were collected using the Leginon data collection software⁴⁴ on a Gatan 4k \times 4k camera using low-dose procedures (20 e⁻ Å⁻² exposures) and a range of defocus values (from -0.5 to -1.2 μ m). Between 300 and 600 images were acquired for each of the negative stain data sets.

Preparation of PIC samples for cryo-EM observation was carried out using 400 mesh C-flats containing 4 μ m holes with 4 μ m spacing (Protochips). A thin carbon film was floated onto the grid before it was plasma cleaned for 5 s using a Solarus plasma cleaner (Gatan) equipped with 75% argon/25% oxygen gas immediately before sample deposition. An aliquot (3 μ L) of the purified sample (~100 nM) was placed onto the grid and loaded into a Vitrobot (FEI) at 100% humidity and 4 °C. The sample was allowed to absorb for 5 min (under low illumination conditions), then was blotted for 4 s and immediately plunged into liquid ethane. The frozen grids were stored in liquid nitrogen until loaded into a Tecnai F20 Twin transmission electron microscope operating at 120 keV using a 626 single-tilt cryotransfer system (Gatan). Data were acquired at a nominal magnification of $\times 100,000$ (1.05 Å per pixel) using low-dose procedures (20 e⁻ Å⁻² exposures) and a range of defocus values (from -1.2 to -2.4 μ m). Between 1,500 and 3,200 images for each of the cryo data sets were collected using the MSI-T application of the Leginon data collection software⁴⁴.

Image processing. Negative stain data pre-processing was performed using the Appion processing environment⁴⁵. Particles were automatically selected from the micrographs using a difference of Gaussians (DoG) particle picker⁵³. The contrast transfer function (CTF) of each micrograph was estimated using both the ACE2 and CTFFind programs during data collection^{54,55}, the phases were flipped using CTFFind, and particle stacks were extracted using a box size of 256 \times 256 pixels (except for both the TFIIF containing PIC complex and free TFIIF samples, which use 320 \times 320 pixel boxes) from images whose ACE2 confidence value was greater than 0.8, followed by normalization using the XMIPP program to remove pixels which were above or below 4.5 σ of the mean value⁵⁶. The particle stack was binned by a factor of two and two-dimensional classification was conducted using iterative multivariate statistical analysis and multireference alignment analysis (MSA-MRA) within the IMAGIC software⁵⁷. Class averages containing properly assembled complexes were manually selected and re-extracted to create a new particle stack for reconstruction.

Cryo data processing was performed in a similar manner as the negative stain data. Particle stack was extracted using 384 \times 384 pixel box size from phase-flipped images and binned by a factor of two.

Three-dimensional reconstruction. The cryo-negative staining structure of free Pol II⁵⁸, after low-pass filtering to 60 Å, was used as initial model for reconstruction of all the negatively stained PIC samples. For reconstruction of the TFIIF containing PIC, which has a substantial extra mass with respect to Pol II, the negative stain reconstruction of the TBP–TFIIA–TFIIB–DNA–Pol II–TFIIF–TFIIE sample, after low-pass filtering to 60 Å, was instead used as the initial reference (Supplementary Fig. 1). For reconstruction of the free TFIIF, the core TFIIF density segmented from the negative stain TBP–TFIIA–TFIIB–DNA–Pol II–TFIIF–TFIIE–TFIIF refined model was used as the initial reference, after low-pass filtering to 60 Å resolution. Three-dimensional reconstruction was conducted using an iterative multi-reference projection-matching approach containing libraries from the EMAN2 and SPARX software packages^{46,47} with two identical copies of the initial model as references. This step allowed us to further eliminate contamination, aggregated, or damaged complexes, which became enriched in one of the reconstructions. Refinement began at an angular step of 25° and progressed down to 4° angular increments. At each step, refinement proceeded to the next angular step only once >95% of the particles had a pixel error of <1 pixel. The particle numbers contributing to the final negative stain reconstructions were 11,880 for TBP–TFIIA–TFIIB–DNA–Pol II, 13,770 for as previous plus TFIIF, 15,656 for as previous plus TFIIE, 64,712 for as previous plus TFIIF, and 13,023 for free TFIIF. The resolution of the reconstruction was

estimated using the 0.143 Fourier shell correlation (FSC) criterion to be about 15 Å for PIC complexes and 20 Å for free TFIIF.

Cryo-EM reconstructions were performed in a similar manner. The negative stain reconstruction of TBP-TFIIA-TFIIB-DNA-Pol II-TFIIF was used as the initial reference for all the cryo reconstructions, except for that of TBP-TFIIA-TFIIB-DNA-Pol II, which used its corresponding negative stain model as the initial reference (Supplementary Figs 2–5). All initial models were low-pass filtered to 60 Å resolution. The particle numbers contributing to the final reconstructions were 122,480 for TBP-TFIIA-TFIIB-DNA-Pol II, 43,785 for as previous plus TFIIF, 51,043 for as previous plus TFIIE, and 53,505 for as previous in the open conformation. To dampen low-resolution amplitudes of the final maps, the Fourier amplitudes were adjusted to match an experimental GroEL SAXS curve using the SPIDER software⁵⁹. The estimated resolution, using the 0.143 FSC criterion, was between 11 and 13 Å for the cryo-EM reconstructions of PIC assembly intermediates (Supplementary Figs 2–5).

Local resolution calculation was performed for all reconstructions using the 'bloccres' function in the Bsoft package^{60,61} (Supplementary Figs 2–6). Volume segmentation, automatic rigid-body docking, figure and movie generation were performed using UCSF Chimera⁴⁸. The globally bent DNA model was generated using the 3D-DART online server⁶² and the 3DNA software package⁶³.

51. Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* **23**, 439–451 (2009).
52. Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005).
53. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
54. Mallick, S. P., Carragher, B., Potter, C. S. & Kriegman, D. J. ACE: automated CTF estimation. *Ultramicroscopy* **104**, 8–29 (2005).
55. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
56. Sorzano, C. O. et al. XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204 (2004).
57. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
58. Kostek, S. A. et al. Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* **14**, 1691–1700 (2006).
59. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
60. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
61. Lander, G. C. et al. Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**, 186–191 (2012).
62. van Dijk, M. & Bonvin, A. M. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* **37**, W235–W239 (2009).
63. Lu, X. J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols* **3**, 1213–1227 (2008).

Carbon monoxide in clouds at low metallicity in the dwarf irregular galaxy WLM

Bruce G. Elmegreen¹, Monica Rubio², Deidre A. Hunter³, Celia Verdugo², Elias Brinks⁴ & Andreas Schruba⁵

Carbon monoxide (CO) is the primary tracer for interstellar clouds where stars form, but it has never been detected in galaxies in which the oxygen abundance relative to hydrogen is less than 20 per cent of that of the Sun, even though such 'low-metallicity' galaxies often form stars. This raises the question of whether stars can form in dense gas without molecules, cooling to the required near-zero temperatures by atomic transitions and dust radiation rather than by molecular line emission¹; and it highlights uncertainties about star formation in the early Universe, when the metallicity was generally low. Here we report the detection of CO in two regions of a local dwarf irregular galaxy, WLM, where the metallicity is 13 per cent of the solar value^{2,3}. We use new submillimetre observations and archival far-infrared observations to estimate the cloud masses, which are both slightly greater than 100,000 solar masses. The clouds have produced stars at a rate per molecule equal to 10 per cent of that in the local Orion nebula cloud. The CO fraction of the molecular gas is also low, about 3 per cent of the Milky Way value. These results suggest that in small galaxies both star-forming cores and CO molecules become increasingly rare in molecular hydrogen clouds as the metallicity decreases.

Wolf–Lundmark–Melotte (WLM) is an isolated dwarf galaxy at the edge of the Local Group⁴. It has a low star-formation rate because of its small size and, like other dwarf irregular (dIrr) galaxies, shows no previous evidence⁵ for the molecular gas that always accompanies young stars in larger galaxies⁶. One problem with the detection of molecules is that the dominant tracer of such gas is CO, and dIrr galaxies have low carbon and oxygen abundances relative to hydrogen. No galaxy with an O/H abundance less than 20% has been detected using CO as a tracer^{7–9}. Far more abundant is molecular hydrogen (H₂), but this does not have an observable state of excitation at the low temperatures (~10–30 K) required for star formation.

To search for star-forming gas, we surveyed WLM for CO($J = 3-2$) emission in rotational state J and for continuum dust emission at 345 GHz using the Atacama Pathfinder Experiment (APEX) telescope at Llano de Chajnantor, Chile, with the Swedish Heterodyne Facility Instrument¹⁰ and the Large APEX Bolometer Camera¹¹ (LABOCA). We also used a map of dust emission at 160 μm from the Spitzer Local Volume Legacy Survey¹² and a map of atomic hydrogen re-reduced from the archives of the Jansky Very Large Array radio telescope. The

dust measurements can be converted to a dust temperature and a dust mass, and, after applying a suitable gas-to-dust ratio, to a gas mass from which the H I mass can be subtracted to give the H₂ mass for comparison with CO.

Figure 1 shows WLM and the two regions, designated A and B, where we detected CO(3–2) emission, along with H I, far-infrared (FIR) and submillimetre images. Observed and derived parameters are listed in Tables 1 and 2, respectively. The peak CO brightness temperature in each detected region is ~0.01–0.015 K and the line-width is ~12 km s^{–1} (full-width at half-maximum). Previous efforts to detect CO($J = 1-0$) in WLM⁵ partly overlapped region A with a 45'' aperture and determined a 5 σ upper limit to the CO(1–0) intensity of 0.18 K km s^{–1}. Our observation with an 18'' aperture yields an intensity of 0.200 ± 0.046 K km s^{–1} for CO(3–2) in the same region. The difference arises because the CO cloud is unresolved even by our 18'' beam—we did not detect comparable CO(3–2) intensities in our searches adjacent to region A. The previous upper limit corresponds to a maximum CO(1–0) luminosity of 8,300 K km s^{–1} pc² inside 45'' (which corresponds to a beam diameter of 215 pc at WLM), whereas the cloud we detect has a CO(3–2) luminosity ~6 times smaller (1,500 K km s^{–1} pc²). Likewise, the previous null detection⁵ in CO($J = 2-1$) claimed a 5 σ upper limit that is about the same as our CO(3–2) detection, but their closest pointing differed from region A by ~70 pc (14'', or half the beam diameter for CO(2–1)), which could have been enough to take it off the CO cloud.

The 160- μm , 870- μm and H I peaks are slightly offset from the CO positions, indicating variations in temperature and molecular fraction. A large H I and FIR cloud that surrounds region A, designated region A1, was used to measure the dust temperature, $T_d \sim 15$ K, which was assumed to be the same throughout the region (the 160- μm observation does not resolve region A, and so a more localized temperature measurement is not possible). We determined T_d from the 870- μm and 160- μm fluxes corrected for the CO(3–2) line and broadband free-free emission (Table 1), assuming a modified black-body function with dust emissivity proportional to frequency to the power β . Local measurements¹³ suggest that $\beta = 1.78 \pm 0.08$, although a range is possible^{14,15}, depending on grain temperature and properties¹⁶. The 870- μm flux was also corrected for an unexplained FIR and submillimetre excess that is commonly observed in other low-metallicity galaxies^{17,18}. An alternate

Table 1 | Observations of WLM

Source	Region	Right ascension	Declination	Beam diameter (")	Flux
CO(3–2)	A	0 h 1 min 57.32 s	–15° 26' 49.5''	18	0.200 ± 0.046 K km s ^{–1}
H I	A	0 h 1 min 57.32 s	–15° 26' 49.5''	22	774 ± 40 mJy km s ^{–1}
870 μm	A	0 h 1 min 57.32 s	–15° 26' 49.5''	22	2.66 ± 0.53 mJy (0.11, 0.02)*
H I	A1	0 h 1 min 56.93 s	–15° 26' 40.84''	45	$4,170 \pm 82$ mJy km s ^{–1}
870 μm	A1	0 h 1 min 56.93 s	–15° 26' 40.84''	45	15.2 ± 3.0 mJy (0.11, 0.06)*
160 μm	A1	0 h 1 min 56.93 s	–15° 26' 40.84''	45	136.2 ± 13.6 mJy (0.05)†
CO(3–2)	B	0 h 2 min 1.68 s	–15° 27' 52.5''	18	0.129 ± 0.032 K km s ^{–1}

* Quantities in parentheses are the CO(3–2) flux and the free-free emission, both in mJy, that were subtracted from the source flux before calculating the dust flux.

† Quantity in parentheses is the free-free emission, in mJy, that was subtracted from the source flux before calculating the dust flux. The average FIR excess factor¹⁸ for the Small Magellanic Cloud (SMC) is 1.7, so we divide the CO-corrected and free-free-corrected 870- μm fluxes in the table by 1.7 to get the thermal dust flux.

¹IBM Research Division, T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, New York 10598, USA. ²Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile.

³Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, Arizona 86001, USA. ⁴Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK. ⁵Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, California 91125, USA.

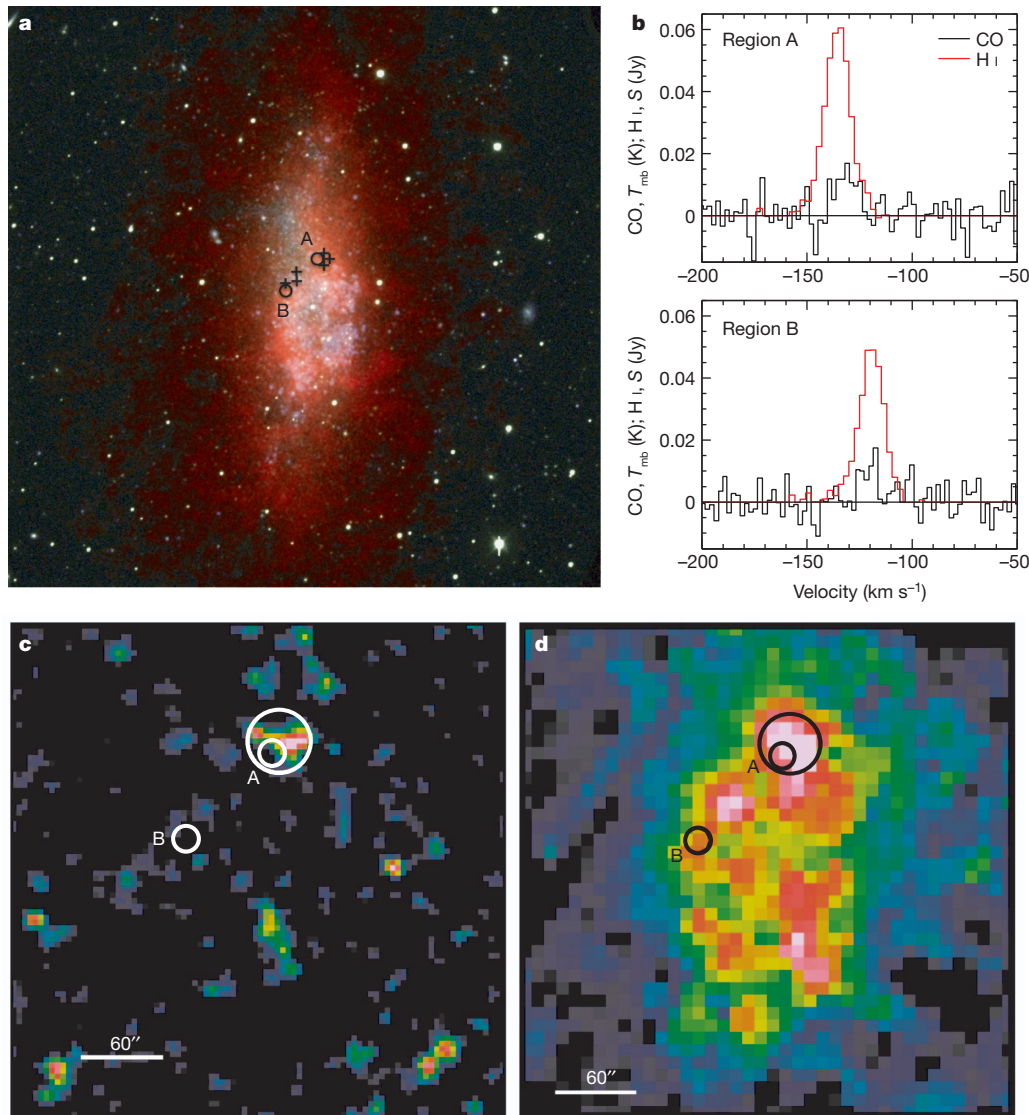


Figure 1 | Observations of the galaxy WLM. WLM is a small, gas-rich galaxy 985 ± 33 kpc from the Milky Way⁴. It contains $1.6 \times 10^7 M_{\odot}$ of stars²⁸, compared with $(6.4 \pm 0.6) \times 10^{10} M_{\odot}$ in the Milky Way²⁹, and it forms new stars at a rate²³ of $0.006 M_{\odot} \text{ yr}^{-1}$, which is 12 times higher per unit stellar mass than the Milky Way³⁰. **a**, Colour composite image: red, H I; green, V band; blue, GALEX far-ultraviolet. For H I the aperture was $7.6''$ and the resolution was 2.6 km s^{-1} , and for CO(3–2) the aperture was $18''$ (circles) and the resolution was 0.11 km s^{-1} , although the CO(3–2) spectra shown in the figure were smoothed to a resolution of 2.2 km s^{-1} . The CO detections are labelled; their exposure times were 218 min (region A) and 248 min (region B). Other regions searched with exposure times shorter by factors of ~ 2 to ~ 6 are indicated by plus signs; the presence of comparable CO mass in some of these other regions cannot be ruled out. **b**, Spectra of the two detections: velocities are relative to the local standard of rest; CO labels main-beam brightness temperature, T_{mb} , in kelvin; H I labels flux in Jy. **c**, False-colour image of the 870- μm observations made with LABOCA on APEX. **d**, False-colour Spitzer 160- μm image obtained from Spitzer archives. In **c** and **d**, the images show the same field of view and the small circles ($22''$ diameter, the resolution of LABOCA) show where CO was detected. The large circle is $45''$ in diameter and surrounds a large H I and FIR cloud (region A1) where the dust temperature was measured.

Table 2 | Derived quantities for WLM

Source	Region	T (K)	Σ^* ($M_{\odot} \text{ pc}^{-2}$)	Mass (M_{\odot})
$\beta = 1.8, \alpha_{\text{CO}} = 124 \pm 60^{\dagger}$				
Dust	A	$14.7 \pm 0.7^{\ddagger}$	0.053 ± 0.014	$(4.6 \pm 1.2) \times 10^2$
Gas §	A		$(58 \pm 15)\delta_{\text{GD}}$	$((5.1 \pm 1.3) \times 10^5)\delta_{\text{GD}}$
H II	A		27.3 ± 1.4	$(2.4 \pm 0.1) \times 10^5$
H ₂	A		31 ± 15	$(1.8 \pm 0.8) \times 10^5$
H ₂ ¶	B		20 ± 10	$(1.2 \pm 0.6) \times 10^5$
$\beta = 1.6, \alpha_{\text{CO}} = 34 \pm 34^{\dagger}$				
Dust	A	$15.9 \pm 0.8^{\ddagger}$	0.032 ± 0.008	$(2.8 \pm 0.7) \times 10^2$
Gas §	A		$(36 \pm 9)\delta_{\text{GD}}$	$((3.1 \pm 0.8) \times 10^5)\delta_{\text{GD}}$
H ₂	A		8.3 ± 9	$(0.5 \pm 0.5) \times 10^5$
H ₂ ¶	B		5.3 ± 6	$(0.3 \pm 0.3) \times 10^5$
$\beta = 2, \alpha_{\text{CO}} = 271 \pm 97^{\dagger}$				
Dust	A	$13.6 \pm 0.6^{\ddagger}$	0.087 ± 0.022	$(7.5 \pm 1.9) \times 10^2$
Gas §	A		$(95 \pm 24)\delta_{\text{GD}}$	$((8.3 \pm 2.1) \times 10^5)\delta_{\text{GD}}$
H ₂	A		67 ± 24	$(3.9 \pm 1.4) \times 10^5$
H ₂ ¶	B		44 ± 16	$(2.5 \pm 0.9) \times 10^5$

* Mass column density of gas or dust.

† The units of α_{CO} are $M_{\odot} \text{ pc}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$. The uncertainty is dominated by the uncertainties in the 160- μm and 870- μm fluxes, as indicated by their error limits in Table 1. The error limits are approximately symmetric.

‡ The dust temperature in region A is assumed to be the same as the measured dust temperature in region A1.

§ $\delta_{\text{GD}} = R_{\text{GD}}/1,100$ is the gas-to-dust ratio, R_{GD} , normalized by the solar value and scaled to the metallicity of WLM. Lowering δ_{GD} lowers α_{CO} , but this does not seem reasonable: data suggests that the gas-to-dust mass ratio is $\sim 5,000$ for $12 + \log(\text{O}/\text{H}) = 7.8$ (ref. 27), and this implies larger values of δ_{GD} (~ 4.5) and α_{CO} . The gas mass and resulting α_{CO} value also depend on the assumed correction factor of 1.7 for submillimetre excess. With no correction for this excess, α_{CO} increases for all β values: at $\beta = 1.8$, $\alpha_{\text{CO}} = 370$. Solutions with no submillimetre excess correction and lower β values¹⁹ can be found in Supplementary Information. In addition, α_{CO} depends on the assumed value of CO(3–2)/CO(1–0), which was taken to be 0.8 in Table 2; a value of CO(3–2)/CO(1–0) = 1 increases α_{CO} to 155 for $\beta = 1.8$.

¶ The H I mass column density is corrected for helium and heavy elements.

‡ The molecular mass for region B was calculated using the CO integrated intensity and the value of α_{CO} determined from region A.

combination of lower β with no submillimetre correction¹⁹ gives similar results (Supplementary Information). The dust mass was calculated from an emissivity¹⁴ of $\kappa = 13.9 \text{ cm}^2 \text{ gm}^{-1}$ at $140 \mu\text{m}$, and converted to $870 \mu\text{m}$ with the same power-law index, β . The dust mass for region A is then $M_{\text{dust,A}} = S_{870,A} D^2 / \kappa B_{\nu}(T_d)$ for flux $S_{870,A}$, distance D and black-body spectral function B_{ν} .

Dust mass is converted to gas mass using a factor equal to the gas-to-dust mass ratio, R_{GD} . An approximation⁸ is to assume the solar value²⁰ ($1/0.007$) increased by the inverse of the metallicity of WLM (0.13), which would give $1,100$. We use this approximation here, but introduce a scaling factor to the gas mass, $\delta_{\text{GD}} = R_{\text{GD}}/1,100$; that is, the gas-to-dust ratio normalized by the solar value and scaled to the metallicity.

The results are in Table 2, assuming $\beta = 1.8$ as the fiducial value and comparing the results with those for $\beta = 1.6$ and 2 to illustrate the dependence of the results on β . Dust and gas mass correlate¹⁵ with the assumed β value. The total gas mass column density in a $22''$ region around region A is $\sim 58 \pm 15 M_{\odot} \text{ pc}^{-2}$ for $\beta = 1.8$ (M_{\odot} , solar mass). Atomic hydrogen contributes $\sim 27.3 \pm 1.4 M_{\odot} \text{ pc}^{-2}$, and the remainder is ascribed to molecular H_2 traced by the observed CO.

The integral under the CO(3–2) line from region A is $I_{\text{CO}} = 0.200 \pm 0.046 \text{ K km s}^{-1}$. This intensity has to be converted to CO(1–0) before comparing it with H_2 mass in the conventional way. We take as a guide²¹ the value of CO(3–2)/CO(1–0) ≈ 0.80 in another low-metallicity galaxy, the SMC²² (where O/H = 20% of the solar value). The result, 0.25 K km s^{-1} , is combined with the H_2 mass column density to determine the conversion factor, α_{CO} , from CO(1–0) to H_2 . If α_{CO} can be calibrated as a function of metallicity, then CO observations can be used directly to infer the molecular gas content irrespective of the dust spectral energy distribution. Extensive compilations^{7–9} show α_{CO} increasing strongly at lower metallicity, from $\sim 4 M_{\odot} \text{ pc}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ in the Milky Way⁸, where the metallicity³ is $12 + \log(\text{O}/\text{H}) = 8.69$, down to the previous CO detection limit⁸ in the SMC, where $\alpha_{\text{CO}} \approx 70 M_{\odot} \text{ pc}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ at $12 + \log(\text{O}/\text{H}) = 8.0$. Our observations of WLM² at a metallicity of $12 + \log(\text{O}/\text{H}) = 7.8$ continue this trend.

Taking the H_2 column density from the residual between the dust-derived total and the H I column density, $31 \pm 15 M_{\odot} \text{ pc}^{-2}$, and dividing by the inferred CO(1–0) line integral of 0.25 K km s^{-1} , we obtain $\alpha_{\text{CO}} = 124 \pm 60 M_{\odot} \text{ pc}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ including helium and heavy elements, with a range in α_{CO} from 34 to 271 as β varies from 1.6 to 2. The corresponding factor, X_{CO} , for conversion from I_{CO} to H_2 column density would be $(5.8 \pm 2.8) \times 10^{21} \text{ cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$, ranging from 1.5×10^{21} to 1.3×10^{22} as β varies between 1.6 and 2. There is a large uncertainty because of the unknown dust properties (β , κ , δ_{GD} and the submillimetre excess) and molecular excitation (CO(3–2)/CO(1–0)) in dIrr galaxies.

The star-formation rate based on the H α and far-ultraviolet²³ fluxes within an $18''$ aperture centred on cloud A is $(3.9\text{--}4.8) \times 10^{-5} M_{\odot} \text{ yr}^{-1}$. Dividing these rates into the CO-associated molecular mass using $\alpha_{\text{CO}} = 124 M_{\odot} \text{ pc}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ gives a CO molecular consumption time (for converting gas into stars) of 4.6–3.8 Gyr for region A. In region B, the star-formation rate from H α and far-ultraviolet fluxes is $(1.7\text{--}12.6) \times 10^{-5} M_{\odot} \text{ yr}^{-1}$ and the CO molecular consumption time is 6.7–1.5 Gyr. These times are only slightly larger than the average value in spiral galaxies²⁴, which is ~ 2 Gyr, but they are ten times larger than the rate per molecule in local giant molecular clouds²⁵, which is a more direct analogy with our observations.

The detection of CO in WLM suggests that star formation continues to occur in dense molecular gas even at lower metallicities than previously observed. The similarity between the metallicities of dIrr galaxies such as WLM and those of larger galaxies at high redshift²⁶ implies that we should be able to study star formation in young galaxies using the usual techniques.

Received 24 October 2012; accepted 23 January 2013.

1. Krumholz, M. R. Star formation in atomic gas. *Astrophys. J.* **759**, 9 (2012).

2. Lee, H., Skillman, E. D. & Venn, K. A. Investigating the possible anomaly between nebular and stellar oxygen abundances in the dwarf irregular galaxy WLM. *Astrophys. J.* **620**, 223–237 (2005).
3. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
4. Leaman, R. et al. The resolved structure and dynamics of an isolated dwarf galaxy: a VLT and Keck spectroscopic survey of WLM. *Astrophys. J.* **750**, 33 (2012).
5. Taylor, C. L. & Klein, U. A search for CO in the Local Group dwarf irregular galaxy WLM. *Astron. Astrophys.* **366**, 811–816 (2001).
6. Bigiel, F. et al. A constant molecular gas depletion time in nearby disk galaxies. *Astrophys. J.* **730**, L13 (2011).
7. Taylor, C. L., Kobulnicky, H. A. & Skillman, E. D. CO emission in low-luminosity, H I-rich galaxies. *Astron. J.* **116**, 2746–2756 (1998).
8. Leroy, A. K. et al. The CO-to- H_2 conversion factor from infrared dust emission across the Local Group. *Astrophys. J.* **737**, 12 (2011).
9. Schrubba, A. et al. Low CO luminosities in dwarf galaxies. *Astron. J.* **143**, 138 (2012).
10. Vassilev, V. et al. A Swedish heterodyne facility instrument for the APEX telescope. *Astron. Astrophys.* **490**, 1157–1163 (2008).
11. Siringo, G. et al. The Large APEX Bolometer Camera LABOCA. *Astron. Astrophys.* **497**, 945–962 (2009).
12. Dale, D. A. et al. The Spitzer Local Volume Legacy: survey description and infrared photometry. *Astrophys. J.* **703**, 517–556 (2009).
13. Planck Collaboration et al. Planck early results. XXV. Thermal dust in nearby molecular clouds. *Astron. Astrophys.* **536**, A25 (2011).
14. Draine, B. T. Interstellar dust grains. *Annu. Rev. Astron. Astrophys.* **41**, 241–289 (2003).
15. Galametz, M. et al. Mapping the cold dust temperatures and masses of nearby KINGFISH galaxies with Herschel. *Mon. Not. R. Astron. Soc.* **425**, 763–787 (2012).
16. Coupeaud, et al. Low-temperature FIR and submillimetre mass absorption coefficient of interstellar silicate dust analogues. *Astron. Astrophys.* **535**, A124 (2011).
17. Galametz, M. et al. Probing the dust properties of galaxies up to submillimetre wavelengths. II. Dust-to-gas mass ratio trends with metallicity and the submm excess in dwarf galaxies. *Astron. Astrophys.* **532**, A56 (2011).
18. Verdugo, C. *Sub-Millimeter Studies of Cold Dust and Gas in the Magellanic Clouds*. MSc thesis, Univ. Chile (2012).
19. Planck Collaboration. Planck early results. XVII. Origin of the submillimetre excess dust emission in the Magellanic Clouds. *Astron. Astrophys.* **536**, A17 (2011).
20. Draine, B. T. et al. Dust masses, PAH abundances, and starlight intensities in the SINGS galaxy sample. *Astrophys. J.* **663**, 866–894 (2007).
21. Nikolić, S., Garay, G., Rubio, M. & Johansson, L. E. B. CO and CS in the Magellanic Clouds: a χ^2 -analysis of multitransitional data based on the MEP radiative transfer model. *Astron. Astrophys.* **471**, 561–571 (2007).
22. Dufour, R. J. The composition of H II regions in the Magellanic Clouds. *IAU Symp.* **108**, 353–360 (1984).
23. Hunter, D. A., Elmegreen, B. G. & Ludka, B. C. GALEX ultraviolet imaging of dwarf galaxies and star formation rates. *Astron. J.* **139**, 447–475 (2010).
24. Leroy, A. K. et al. The star formation efficiency in nearby galaxies: measuring where gas forms stars effectively. *Astron. J.* **136**, 2782–2845 (2008).
25. Lada, C. J., Forbrich, J., Lombardi, M. & Alves, J. F. Star formation rates in molecular clouds and the nature of the extragalactic scaling relations. *Astrophys. J.* **745**, 190 (2012).
26. Mannucci, F. et al. LSD: Lyman-break galaxies, stellar populations and dynamics – I. Mass, metallicity and gas at $z \sim 3.1$. *Mon. Not. R. Astron. Soc.* **398**, 1915–1931 (2009).
27. Engelbracht, C. W. et al. Metallicity effects on dust properties in starbursting galaxies. *Astrophys. J.* **678**, 804–827 (2008).
28. Zhang, H.-X., Hunter, D. A., Elmegreen, B. G., Gao, Y. & Schrubba, A. Outside-in shrinking of the star-forming disks of dwarf irregular galaxies. *Astron. J.* **143**, 47 (2012).
29. McMillan, P. J. Mass models of the Milky Way. *Mon. Not. R. Astron. Soc.* **414**, 2446–2457 (2011).
30. Chomiuk, L. & Povich, M. S. Toward a unification of star formation rate determinations in the Milky Way and other galaxies. *Astron. J.* **142**, 197 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded in part by the US National Science Foundation through grants AST-0707563 and AST-0707426 to D.A.H. and B.G.E. M.R. and C.V. wish to acknowledge support from CONICYT (FONDECYT grant no. 1080335). M.R. was also supported by the Chilean Center for Astrophysics FONDAF grant no. 15010003. A.S. was supported by the Deutsche Forschungsgemeinschaft Priority Program 1177. We are grateful to M. Albrecht for help with the LABOCA data reduction and to L. Hill for making Fig. 1a. The National Radio Astronomy Observatory is a facility of the US National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Author Contributions B.G.E. coordinated the observational team, did the calculations for Table 2 and wrote the manuscript; M.R. was principal investigator for Chilean observing time on the APEX telescope and, with C.V., observed the galaxy in CO and at $870 \mu\text{m}$, reduced the relevant data in Table 1 and did relevant calculations for Table 2; D.A.H. determined the observational strategy, selected WLM for study, chose the observing coordinates, extracted the H I spectra from the LITTLE THINGS data and prepared Fig. 1. E.B. was principal investigator on the APEX proposal using European time through ESO and coordinated the work on data uncertainties and background noise. A.S. made the WLM H I data cube from Jansky Very Large Array observations. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to and requests for materials should be addressed to B.G.E. (bge@us.ibm.com).

Evidence for sympathetic vibrational cooling of translationally cold molecules

Wade G. Rellergert¹, Scott T. Sullivan¹, Steven J. Schowalter¹, Svetlana Kotochigova², Kuang Chen¹ & Eric R. Hudson¹

Compared with atoms, molecules have a rich internal structure that offers many opportunities for technological and scientific advancement. The study of this structure could yield critical insights into quantum chemistry^{1–3}, new methods for manipulating quantum information^{4,5}, and improved tests of discrete symmetry violation^{6,7} and fundamental constant variation^{8–10}. Harnessing this potential typically requires the preparation of cold molecules in their quantum rovibrational ground state. However, the molecular internal structure severely complicates efforts to produce such samples. Removal of energy stored in long-lived vibrational levels is particularly problematic because optical transitions between vibrational levels are not governed by strict selection rules, which makes laser cooling difficult. Additionally, traditional collisional, or sympathetic, cooling methods are inefficient at quenching molecular vibrational motion¹¹. Here we experimentally demonstrate that the vibrational motion of trapped BaCl^+ molecules is quenched by collisions with ultracold calcium atoms at a rate comparable to the classical scattering, or Langevin, rate. This is over four orders of magnitude more efficient than traditional sympathetic cooling schemes¹¹. The high cooling rate, a consequence of a strong interaction potential (due to the high polarizability of calcium), along with the low collision energies involved¹², leads to molecular samples with a vibrational ground-state occupancy of at least 90 per cent. Our demonstration uses a novel thermometry technique that relies on relative photodissociation yields. Although the decrease in vibrational temperature is modest, with straightforward improvements it should be possible to produce molecular samples with a vibrational ground-state occupancy greater than 99 per cent in less than 100 milliseconds. Because sympathetic cooling of molecular rotational motion is much more efficient than vibrational cooling in traditional systems, we expect that the method also allows efficient cooling of the rotational motion of the molecules. Moreover, the technique should work for many different combinations of ultracold atoms and molecules.

Precision control over the quantum states of atoms has allowed tremendous advances in both applied and fundamental physics, ranging from ultraprecise clocks¹³ and gravity gradiometers¹⁴, to stringent tests on discrete symmetry violations¹⁵ and the production of exotic phases of matter such as degenerate Bose and Fermi gases¹⁶. Extension of these studies to the next simplest such physical system, that is, a diatomic molecule, promises even further progress. Typically, the starting point for these studies is the production of an ultracold sample, which in the case of atoms is usually attained by laser cooling. Applying this technique to molecules, however, is technically challenging, owing mainly to the molecular vibrational degree of freedom, which has no associated angular momentum. Thus, unlike transitions between different rotational and electronic levels, transitions between vibrational levels of a molecule are not governed by strict selection rules. As a result, in general, a molecule simply cannot scatter the necessary number of photons needed to enter the ultracold regime before spontaneously decaying to long-lived vibrational levels that are out of resonance with the cooling lasers. Although it does seem that a subset of molecules

are potentially amenable to laser cooling, as has been recently demonstrated for SrF (ref. 17) and YO (ref. 18), a more general cooling technique is highly desirable. One promising, general method is to use a sample of laser-cooled ultracold atoms to collisionally, that is, sympathetically, cool a co-trapped sample of molecules to ultralow temperatures. Particularly attractive for this technique are charged molecules, which, in contrast to neutral molecules, can be trapped over a broad temperature range for extended periods of time in radio-frequency ion traps, allowing ample time for sympathetic cooling to occur.

So far, sympathetic cooling of charged molecules has been demonstrated with both laser-cooled atomic ions and cryogenic buffer gases¹¹. Unfortunately, neither of these methods can produce molecules that are simultaneously ultracold in both their internal and external degrees of freedom. On the one hand, sympathetic cooling with co-trapped, laser-cooled atomic ions quickly cools molecular ions to very low translational temperatures, but, owing to the long-range nature of the Coulomb interaction, the method yields little internal-state relaxation of the molecules at low collision energies¹⁹. To overcome this limitation, techniques of optical pumping^{20,21} and state-selective photoionization of neutral molecules²² have been used in tandem with atomic ion sympathetic cooling to produce translationally cold molecular ions in the lowest few rotational states. Despite their impressive results, these techniques lack generality because they are restricted to certain classes of molecule and require molecule-specific lasers. On the other hand, sympathetic cooling with helium buffer gas has been demonstrated to yield cooling of both the translational and internal states of molecules owing to the short-range nature of the ion-neutral collision; however, the method can only be used at temperatures above ~ 300 mK, the point at which the vapour pressure of ^3He becomes negligibly small. Additionally, it is relatively inefficient as a result of the low polarizability of the helium atoms.

An alternative to these methods is sympathetic cooling with ultracold, laser-cooled atoms, which both allows access to submillikelvin translational temperatures and is predicted to be very efficient at cooling the internal degrees of freedom because their polarizabilities are more than 100 times greater than that of helium atoms. Although this scheme uses proven technologies that have existed for decades, it has not previously been implemented, possibly owing to the misconception that molecular ions cannot undergo primarily elastic collisions with neutral atoms that are amenable to laser cooling, but instead undergo charge-exchange chemical reactions leading to energetic, neutral molecules. Previously, we have shown that, on the contrary, many combinations of molecular ions and neutral laser-cooled atoms (including the combination of species used here) can coexist without undergoing a chemical reaction when colliding²³. Here we present evidence of internal-state cooling of trapped molecular ions due to sympathetic cooling collisions with ultracold atoms co-located in a magneto-optical trap (MOT). Although the demonstrated degree of cooling is modest, the method is found to be extremely efficient, with a vibrational quenching rate constant on the order of the classical, or Langevin, atom–ion scattering rate constant. Thus, straightforward

¹Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA. ²Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, USA.

improvements in the apparatus should allow production of molecular samples in the vibrational ground state in less than 100 ms. In what follows, we detail a novel technique that we have developed to measure relative populations of molecular vibrational levels, briefly describe the apparatus and experimental approach used to sympathetically cool BaCl^+ molecules, and present data detailing the measured vibrational sympathetic cooling rate.

To quantify the degree of sympathetic cooling of the molecular internal levels, it is necessary to measure the population distribution of rotational and vibrational levels. Typically, this measurement is accomplished with spectroscopic techniques, such as predissociation or multiphoton ionization/dissociation spectroscopy; however, as for most molecular ions, the detailed spectroscopic information necessary for these techniques is not yet known for BaCl^+ . To overcome this challenge, we develop a new route to demonstrate vibrational spectroscopic thermometry experimentally, which uses broadband molecular photodissociation and as such does not require detailed molecular structure data. This novel and general method is applicable to any molecule that can be directly photodissociated and allows an accurate measurement of the molecular vibrational temperature, as long as the internal-state population can be probed on a timescale shorter than or comparable to the vibrational relaxation time. The method exploits the fact that, although the photodissociation cross-section is broad, the individual vibrational levels have unique frequency responses for photodissociation, which are given by²⁴

$$\sigma_{vj}(v) = 4\pi^2 \alpha a_0^2 \frac{h\nu}{E_h} \frac{1}{2J+1} \sum_{J', M, M'} \frac{|\langle A, \varepsilon J' M' | d_z | X, v J M \rangle|^2}{(ea_0/\sqrt{E_h})^2}$$

where E_h is the Hartree energy, e is the charge of the electron, a_0 is the Bohr radius, α is the fine-structure constant, h is Planck's constant, $h\nu$ is the photon energy, ε is the energy of the continuum wavefunction and v is the vibrational quantum number of the X state. The rovibrational wavefunctions of the ground state, $^1\Sigma^+$, and first electronic excited state, $^1\Pi$ (Fig. 1a), are respectively denoted $|X\rangle$ and $|A\rangle$, with rotational quantum numbers J and J' , which have space-fixed z -axis projections of M and M' . The quantity $\langle A | d_z | X \rangle / (ea_0/\sqrt{E_h})$ is dimensionless and contains both the radial and angular parts of the dipole moment, d . Because the continuum wavefunctions associated with solutions of the $A^1\Pi$ molecular potential are highly oscillatory, the expectation value of the dipole moment operator can be accurately approximated by considering only the nature of $|A\rangle$ near the classical turning point, R_C . Thus, if the repulsive potential is approximated as being linearly dependent on R , with slope dV_A/dR , near $R = R_C$, these wavefunctions are well described by Airy functions and the vibrational photodissociation cross-section is given by

$$\sigma_{vj}(v) = \frac{4\pi^2 \alpha}{e^2} \frac{h\nu}{2J+1} \left(d(R_C) \frac{\phi_v(R_C)}{\sqrt{dV_A/dR}|_{R=R_C}} \right)^2 \quad (1)$$

where ϕ_v are the real-valued vibrational wavefunctions. Because R_C is linearly related to the photon energy, the shape of the photodissociation cross-section directly reflects the squared probability amplitude of the molecular vibrational wavefunction.

Implementing the photodissociation thermometry technique to be discussed requires knowledge of a photodissociation transition from which an accurate determination of the ϕ_v can be obtained. Recently, we identified such a photodissociation transition in BaCl^+ (ref. 24). However, to implement the thermometry technique, it was necessary to measure the shape of the photodissociation cross-section to higher precision. To this end, we have developed a novel, compact time-of-flight apparatus²⁵, which enabled us to measure the photodissociation spectrum of BaCl^+ about 1,000 times more precisely than in ref. 24. From this data (Fig. 1c), slight adjustments to the molecular potentials determined in ref. 24 have been made, and the contributions

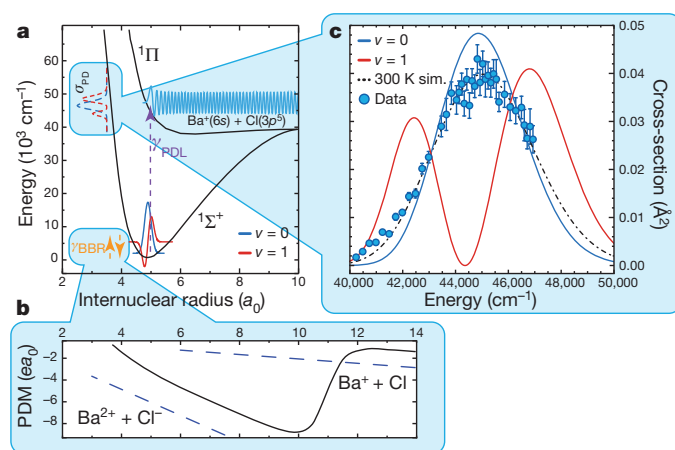


Figure 1 | Photodissociation thermometry of BaCl^+ . **a**, Potential energy curves for the $X^1\Sigma^+$ and $A^1\Pi$ electronic states and a schematic diagram indicating the photodissociation thermometry method described in the text. The different frequency shapes of the $v=0$ and $v=1$ wavefunctions give rise to the different frequency responses of the two levels to the photodissociation laser. γ_{BBR} , redistribution rate due to black-body radiation; γ_{PDL} , transition rate due to the photodissociation laser given the cross-section σ_{PDL} . **b**, The permanent dipole moment (PDM) of BaCl^+ calculated as described in Methods. **c**, Improved data (errors, s.e.) obtained for the cross-section of the $A^1\Pi \leftarrow X^1\Sigma^+$ photodissociation transition in BaCl^+ using a separate ion trap and time-of-flight system. The dash-dot line is the thermally averaged cross-section at 300 K. Also shown are the improved calculations of the individual vibrational-level contributions of the $v=0$ and $v=1$ levels. Although our model (Methods) incorporates the lowest four vibrational levels, here we show only the $v=0$ and $v=1$ levels, for clarity.

of the individual vibrational states to the total photodissociation cross-section can be identified, the lowest two of which (that is, $v=0$ and $v=1$) are shown as solid lines in Fig. 1c.

With the individual vibrational-level cross-sections in hand, the crux of the thermometry technique can be understood as follows. Once equilibration has been established with the vacuum apparatus, which operates at 300 K, $\sim 79\%$ of the molecules are in the $v=0$ level, $\sim 15\%$ are in the $v=1$ level and the remaining population is essentially all in the $v=2$ and $v=3$ levels. Although the model introduced later incorporates all of these levels, for illustrative purposes it is sufficient to consider only the $v=0$ and $v=1$ levels. Because the wavefunction of the $v=1$ level has a zero crossing (Fig. 1a) the photodissociation cross-section contains a node (Fig. 1b), which occurs at $\lambda = 225$ nm ($\nu/c = 44,400$ cm^{-1}). A laser tuned to this wavelength will photodissociate molecules in the $v=0$ level, but will not affect molecules in the $v=1$ level. Furthermore, as can be inferred by thermally averaging equation (1), at the temperatures relevant to this work the rotational distribution does not change the resulting cross-sections. As a result, if collisions with the ultracold Ca atoms quench the vibrational motion of the molecules, this quenching will be signalled by an enhanced photodissociation rate at this wavelength. Conversely, by tuning the photodissociation laser (PDL) closer to one of the maxima of the $v=1$ cross-section, MOT-induced quenching will instead be signalled by suppression of the observed photodissociation rate as molecules are being cooled to the $v=0$ level, lowering the total cross-section. As a result, we can infer the change in vibrational temperature from the change in photodissociation rate at different photodissociation wavelengths.

The hybrid atom-ion apparatus, which we call the MOTion trap, used to measure these effects is similar to the one used in refs 26, 27 (a detailed schematic diagram can be found in Supplementary Information). Typical data obtained to demonstrate vibrational cooling is plotted in Fig. 2a. To begin, BaCl^+ molecules created by laser ablation are trapped in a linear quadrupole ion trap (Fig. 2a, i), which is co-located with a MOT that can be loaded with ultracold Ca atoms. Notably, without any cooling, the initial internal and translational

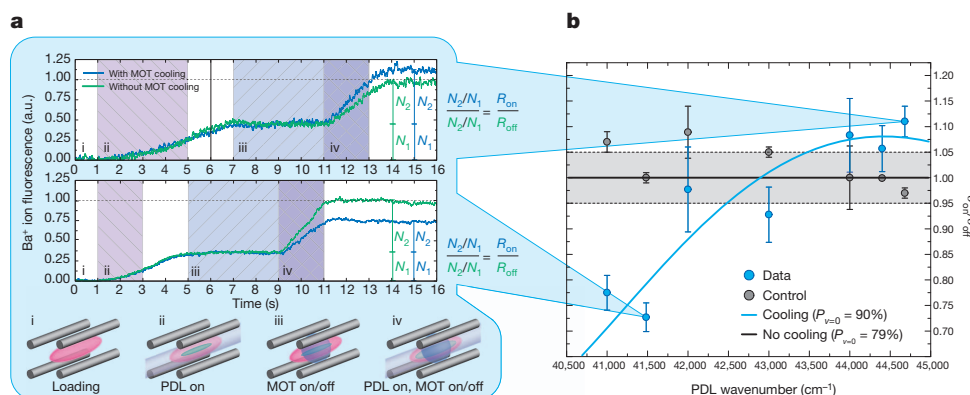


Figure 2 | Measurement of MOT-induced vibrational quenching of BaCl^+ ions. **a**, Example traces of the Ba^+ ion fluorescence measured during the experimental sequence. Initially the trap is loaded with a pure BaCl^+ sample (pink) (i). The PDL is then turned on to create a mixed sample of BaCl^+ and Ba^+ ions (green) (ii). Once the Ba^+ ions have been laser-cooled, the MOT (blue) is optionally turned on and the BaCl^+ ions are collisionally cooled by the ultracold Ca atoms (iii). Following that, the BaCl^+ ions are probed with the PDL to determine the amount of vibrational relaxation due to the MOT (iv). The ratio of the numbers of Ba^+ ions created in each instance of photodissociation with the MOT respectively on or off is then calculated. The

temperatures of the BaCl^+ molecules are determined by different mechanisms and are established in the first few seconds after the trap is loaded. Internally, the BaCl^+ molecules thermalize to 300 K owing to black-body radiation emitted by the room-temperature vacuum apparatus, while translationally the temperature is set by the ion dynamics in the trap. To attain measurements with the high signal-to-noise ratio necessary to demonstrate vibrational cooling, the system requires the use of an ion cloud composed of $\sim 10^3$ BaCl^+ molecules. At these numbers, the initial translational temperature of the BaCl^+ molecules is primarily set to $\sim 1,000$ K by collisional heating through micromotion interruption, which results when collisions between ions disrupt the stable ion trajectories in the trap allowing the radio-frequency confining potential to do net work on the ions²⁸. Therefore, for this experiment, we use additional translational cooling of the BaCl^+ molecules by sympathetic cooling with co-trapped, laser-cooled Ba^+ ions. These Ba^+ ions are loaded into the trap with 100% efficiency by photodissociating a fraction of the BaCl^+ molecules using a pulsed dye laser as discussed in ref. 24 (Fig. 2a, ii), and, once laser-cooled, sympathetically cool the BaCl^+ molecules to millikelvin translational temperatures¹⁹.

These atomic ions also provide an important diagnostic, because their fluorescence is used both as a means of determining overlap of the MOTion trap²⁶ as well as the initial, proportional number of BaCl^+ molecules.

Once the BaCl^+ molecules are translationally cold, as indicated by the steady-state value of the Ba^+ ion fluorescence and camera images of the ion cloud, the MOT is turned on (Fig. 2a, iii) and short-range collisions with the highly polarizable, neutral Ca atoms result in internal-state cooling. To measure the effect of the cooling, the PDL is then turned on once again (Fig. 2a, iv) and the increase in fluorescence yields the number of Ba^+ ions produced, thus allowing an inference of the ground-state vibrational population of the BaCl^+ molecules.

Several sources of systematic error can be eliminated (for example the variable number of BaCl^+ molecules produced in the loading process and possible inaccuracies in the absolute cross-section value, the PDL intensity or both) by taking the ratio of the final number of Ba^+ ions produced to the initial number, $R_{\text{on}} = N_2/N_1$ (Methods and Fig. 2a). We then repeat this measurement sequence without turning on the Ca MOT and determine the same ratio, which we denote R_{off} .

top and bottom panels are data for PDL wavenumbers of 41,500 and 44,675 cm^{-1} , respectively. a.u., arbitrary units. **b**, The average of all data points (errors, s.e.) obtained, as described in Methods, is plotted for various photodissociation wavenumbers. The blue points are cooling data and the grey points are a control analysis that should be consistent with unity (Methods). The curves are the expected result from a rate equation model of the experimental sequence described in Methods. The agreement with the data shows that the BaCl^+ sample has been cooled to a ground-state vibrational population of $\sim 90\%$ in the average case.

It is straightforward to show that in the linear regime in which we operate, the ratio $R_{\text{on}}/R_{\text{off}}$ is equivalent to the ratio of the total cross-sections in each case, that is, $\sigma_{\text{on}}/\sigma_{\text{off}}$. This sequence is repeated several times at several PDL wavenumbers. The analysed data (Methods) are plotted in Fig. 2b. These data are compared with a control analysis at each wavelength, which compares measurements for which the MOT remained on or off for both points, and the computed ratio should thus be consistent with unity (Methods). From this comparison, it is clear that the deviation of the actual data is the result of MOT-induced vibrational relaxation.

To quantify the degree of cooling, we use a simple rate equation model that incorporates both the continual redistribution of the molecules' internal states due to black-body radiation throughout the experiment and the MOT-induced vibrational relaxation when the MOT is present (Methods and Supplementary Information). Because the MOT density and overlap with the ions vary for each of the data points in Fig. 2b, a best-fit line using the rate equation is unsuitable. Rather, using the calculated effective density²⁶ for each data point, along with the rate equation model, we determine an average vibrational quenching rate constant of $k \approx 1 \times 10^{-9} \text{ cm}^3 \text{ s}^{-1}$. Using this rate constant with the typical peak MOT atomic density of $\sim 1 \times 10^9 \text{ cm}^{-3}$ and a typical overlap factor of 0.2 (Methods), we calculate the average expected values of $R_{\text{on}}/R_{\text{off}}$ as predicted from the rate equation model (Fig. 2b, blue curve). Because the data agree well with this curve, we use the rate equations to extract the vibrational ground-state population, which in the case of sympathetic cooling by the Ca MOT is found to be $P_{v=0} \geq 90\%$, set by the competition of sympathetic cooling and black-body radiation heating from the chamber walls. In contrast, if the MOT had provided no cooling, the data would have been consistent with the black line in Fig. 2b, which corresponds to a vibrational ground state population of $\sim 79\%$.

To our knowledge, rigorous quantum scattering calculations of the vibrational quenching rate have never been performed in any similar system. However, the large value of the rate constant, which is more than 10^4 times that observed in traditional sympathetic cooling schemes, is probably due to the strong, long-range C_4/R^4 interaction potential between Ca and BaCl^+ ($C_4 = -\alpha_{\text{Ca}}e^2$, where α_{Ca} is the polarizability of Ca). Classically, particles colliding on this potential with energy E and an impact parameter less than the Langevin length, $b_L = \sqrt{C_4/E}$ undergo violent collision trajectories that rapidly spiral

inward towards $R = 0$. Quantum mechanically speaking, the system forms a three-body collision complex with reduced mass μ at the Langevin rate, $k_L = \pi\sqrt{2C_4/\mu}$, which facilitates redistribution of the original molecule's vibrational energy between all three particles. Thus, when the collision complex predissociates back into its parent constituents, this energy is carried away by the sympathetic cooling partner, leaving the molecule in a lower vibrational state. The fact that the observed rate constant is of the same order as the Langevin rate ($k \approx k_L/5$) supports this interpretation¹² and suggests that sympathetic cooling of molecular internal degrees of freedom will be possible with any collision partner that gives rise to long-range interactions conducive to collision complex formation, for example dipole–dipole and monopole–dipole collisions.

Additionally, we have observed no evidence of any chemical reactions between BaCl^+ and electronically excited Ca atoms in the MOT, even though such reactions are energetically allowed. This supports the arguments of refs 27, 29, which state that these collisions are suppressed because the long-range atom–ion interaction shifts the atom out of resonance with the atomic cooling laser at large atom–ion separations and thus prevents atoms in their electronically excited state from colliding with the molecular ions. This also provides an important simplification to the originally proposed method²³, because it eliminates the need for an optical dipole trap for the ultracold atoms. Furthermore, it implies that the rate at which sympathetic cooling collisions occur can be increased without loss of molecules. Several straightforward experimental modifications, such as modifying the ion trap geometry such that the ion cloud is fully contained inside the Ca MOT and replenishing the Ca atom source, are expected to increase the effective Ca MOT density and, as a result, the sympathetic cooling collision rate by a factor of ~ 250 . Such improvements would produce a vibrational ground-state population greater than 99% in less than 100 ms, and would potentially render the currently used Ba^+ translational sympathetic cooling unnecessary.

Finally, owing to incomplete spectroscopic information on the BaCl^+ molecule, direct measurement of rotational energy quenching is currently not possible. However, because all evidence from sympathetic cooling systems used so far shows that collisional quenching of the rotational energy proceeds at a much higher rate than vibrational quenching, it is likely that molecules have been cooled rotationally as well; experiments are currently underway to prove this. Ultimately, the minimum rovibrational temperature attainable with this method will be set by the ion radial micromotion, which determines the lowest collision energy with the Ca atoms. This collision energy can be estimated from the amplitude of the micromotion, which is given as a function of electric field as $a_{\text{mm}} = \sqrt{2}eE(r)/m\omega\Omega$, where ω and Ω are respectively the secular and drive frequencies³⁰. The average micromotion kinetic energy ($m v_{\text{mm}}^2/2 = m(a_{\text{mm}}\Omega)^2/2$, where m is the ion mass and v_{mm} is the component of the ion velocity due to micromotion) over a typical cloud profile in our apparatus for typical trap parameters (that is, a radial pseudo-potential of depth ~ 10 eV), is ~ 1 K, whereas the maximum possible micromotion contribution to a collision (for ions residing at the edge of the cloud) is ~ 10 K. Because the BaCl^+ vibrational splitting corresponds to ~ 500 K, this energy is inconsequential for vibrational cooling. However, the BaCl^+ rotational splitting is ~ 0.5 K, meaning that in the present experiment the lowest few rotational levels could still be populated. Therefore, future experiments aiming to produce samples of molecules purely in the rovibrational ground state should use smaller samples of molecular ions, so that the cloud size and, thus, the micromotion are reduced to the point where collision energies are below the molecular rotational splitting energy. In the present experiment, we estimate that this regime can be reached by working with a sample of order one-tenth the size of the current cloud.

We have demonstrated a broadly applicable, general technique to cool the vibrational motion of polar molecules. We have also developed a general technique, photodissociation thermometry, for

probing molecular vibrational states, which we have used to determine the vibrational quenching rate constant and the vibrational ground-state population. This result opens a new route to the production of ultracold molecules and is expected to enable myriad experiments in quantum chemistry^{1–3}, precision tests of fundamental physics³¹, and quantum information and simulation^{4,5}.

METHODS SUMMARY

The data acquisition system takes an alternating sequence of data points for which the MOT is either on or off during the measurement, and acquires over 100 data points at each PDL wavelength. For each wavelength, the ratio of successive points ($R_{\text{on}}/R_{\text{off}}$) is computed to eliminate unidentified time varying systematic errors, and the results are averaged. For the control analysis, the ratios between successive data points of the same type (for example $R_{\text{on},1}/R_{\text{on},2}$, $R_{\text{off},1}/R_{\text{off},2}$ and so on) are instead computed and the results are averaged. See Methods for details.

The production of Ba^+ ions in the experiment is modelled using a simple rate equation model that includes the lowest four vibrational levels of the BaCl^+ molecules and accounts for population redistribution among the levels throughout the experimental sequence. This redistribution of the levels is due to collisions with the ultracold Ca atoms as well as spontaneous and stimulated emission and absorption of photons by the molecules owing to the black-body radiation field. Because knowledge of the dipole moment of the electronic ground state of BaCl^+ is necessary to determine the spontaneous and stimulated emission and absorption rates, we calculated the dipole moment using a non-relativistic, multiconfigurational, second-order perturbation theory (CASPT2) implemented in the MOLCAS software suite. The result is shown in Fig. 1b. See Methods for details.

Full Methods and any associated references are available in the online version of the paper.

Received 9 October 2012; accepted 23 January 2013.

- Smith, D. The ion chemistry of interstellar clouds. *Chem. Rev.* **92**, 1473–1485 (1992).
- Balakrishnan, N. & Dalgarno, A. Chemistry at ultracold temperatures. *Chem. Phys. Lett.* **341**, 652–656 (2001).
- Krems, R. V. Cold controlled chemistry. *Phys. Chem. Chem. Phys.* **10**, 4079–4092 (2008).
- André, A. *et al.* A coherent all-electrical interface between polar molecules and mesoscopic superconducting resonators. *Nature Phys.* **2**, 636–642 (2006).
- Schuster, D., Bishop, L., Chuang, L., DeMille, D. & Schoelkopf, R. Cavity QED in a molecular ion trap. *Phys. Rev. A* **83**, 012311 (2011).
- Kozlov, M. G. & Labowsky, L. Z. Parity violation effects in diatomics. *J. Phys. At. Mol. Opt. Phys.* **28**, 1933–1961 (1995).
- Vutha, A. C. *et al.* Search for the electric dipole moment of the electron with thorium monoxide. *J. Phys. At. Mol. Opt. Phys.* **43**, 074007 (2010).
- DeMille, D. *et al.* Enhanced sensitivity to variation of m_e/m_p in molecular spectra. *Phys. Rev. Lett.* **100**, 043202 (2008).
- Chin, C., Flambaum, V. V. & Kozlov, M. G. Ultracold molecules: new probes on the variation of fundamental constants. *N. J. Phys.* **11**, 055048 (2009).
- Hudson, E. *et al.* Production of cold formaldehyde molecules for study and control of chemical reaction dynamics with hydroxyl radicals. *Phys. Rev. A* **73**, 063404 (2006).
- Krems, R. V., Stwalley, W. C. & Friedrich, B. (eds) *Cold Molecules: Theory, Experiment, Applications* Ch. 13, 18 (CRC, 2009).
- Ferguson, E. E. Vibrational quenching of small molecular ions in neutral collisions. *J. Phys. Chem.* **90**, 731–738 (1986).
- Chou, C. W., Hume, D. B., Koelemeij, J. C. J., Wineland, D. J. & Rosenband, T. Frequency comparison of two high-accuracy Al^+ optical clocks. *Phys. Rev. Lett.* **104**, 070802 (2010).
- McGuirk, J., Foster, G., Fixler, J., Snadden, M. & Kasevich, M. Sensitive absolute-gravity gradiometry using atom interferometry. *Phys. Rev. A* **65**, 033608 (2002).
- Griffith, W. *et al.* Improved limit on the permanent electric dipole moment of ^{199}Hg . *Phys. Rev. Lett.* **102**, 101601 (2009).
- Levin, K., Fetter, A. L. & Stamper-Kurn, D. M. (eds) *Ultracold Bosonic and Fermionic Gases* (Elsevier, 2012).
- Shuman, E. S., Barry, J. F. & Demille, D. Laser cooling of a diatomic molecule. *Nature* **467**, 820–823 (2010).
- Hummon, M., Yeo, M., Stuhl, B., Xia, Y. & Ye, J. Direct laser cooling of yttrium monoxide. *Bull. Am. Phys. Soc.* **57**, abstr. C2.00002 (2012).
- Ostendorf, A. *et al.* Sympathetic cooling of complex molecular ions to millikelvin temperatures. *Phys. Rev. Lett.* **97**, 243005 (2006).
- Staanum, P. F., Højbjerg, K., Skyt, P. S., Hansen, A. K. & Drewsen, M. Rotational laser cooling of vibrationally and translationally cold molecular ions. *Nature Phys.* **6**, 271–274 (2010).
- Schneider, T., Roth, B., Duncker, H., Ernsting, I. & Schiller, S. All-optical preparation of molecular ions in the rovibrational ground state. *Nature Phys.* **6**, 275–278 (2010).

22. Tong, X., Winney, A. & Willitsch, S. Sympathetic cooling of molecular ions in selected rotational and vibrational states produced by threshold photoionization. *Phys. Rev. Lett.* **105**, 143001 (2010).
23. Hudson, E. Method for producing ultracold molecular ions. *Phys. Rev. A* **79**, 032716 (2009).
24. Chen, K. *et al.* Molecular-ion trap-depletion spectroscopy of BaCl^+ . *Phys. Rev. A* **83**, 030501 (2011).
25. Schowalter, S. J., Chen, K., Rellergert, W. G., Sullivan, S. T. & Hudson, E. R. An integrated ion trap and time-of-flight mass spectrometer for chemical and photo-reaction dynamics studies. *Rev. Sci. Instrum.* **83**, 043103 (2012).
26. Rellergert, W. *et al.* Measurement of a large chemical reaction rate between ultracold closed-shell ^{40}Ca atoms and open-shell $^{174}\text{Yb}^+$ ions held in a hybrid atom-ion trap. *Phys. Rev. Lett.* **107**, 243201 (2011).
27. Sullivan, S. T., Rellergert, W. G., Kotochigova, S. & Hudson, E. R. Role of electronic state excitation in ground-state-forbidden inelastic collisions between ultracold atoms and ions. *Phys. Rev. Lett.* **109**, 223002 (2012).
28. DeVoe, R. Power-law distributions for a trapped ion interacting with a classical buffer gas. *Phys. Rev. Lett.* **102**, 063001 (2009).
29. Band, Y. & Julienne, P. Optical-Bloch-equation method for cold-atom collisions: Cs loss from optical traps. *Phys. Rev. A* **46**, 330–343 (1992).
30. Major, F. G. & Dehmelt, H. G. Exchange-collision technique for the rf spectroscopy of stored ions. *Phys. Rev.* **170**, 91–107 (1968).
31. Leanhardt, A. *et al.* High-resolution spectroscopy on trapped molecular ions in rotating electric fields: a new approach for measuring the electron electric dipole moment. *J. Mol. Spectrosc.* **270**, 1–25 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by ARO grant no. W911NF-10-1-0505, US NSF grant nos PHY-1005453 and PHY-1205311, and AFOSR grant no. FA 9550-11-1-0243.

Author Contributions E.R.H., W.G.R. and S.T.S. conceived of the thermometry technique and measurement protocol. W.G.R. and S.T.S. built the MOTion trap apparatus, wrote the data acquisition software, and acquired and analysed all data in Fig. 2. S.J.S. built the time-of-flight apparatus, acquired and analysed the data in Fig. 1, and helped K.C. write the data acquisition software for these data. S.K. calculated the potential energy curves and dipole moment for the BaCl^+ molecules as well as the relevant absorption, spontaneous and stimulated emission rates due to black-body radiation. S.J.S. prepared all of the figures. W.G.R. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.G.R. (wgr6@ucla.edu).

METHODS

To eliminate several sources of systematic error when comparing our data with the rate equation model, for example the variable number of BaCl^+ ions produced in the loading process and possible inaccuracies in the absolute cross-section value, the PDL intensity or both, we use the fluorescence as a measure of the number of Ba^+ ions created in each instance of photodissociation and calculate the ratio of the final amount to the initial amount, $R_{\text{on}} = N_2/N_1$ (Fig. 2a). We then repeat this measurement sequence without turning on the Ca MOT, and record the same ratio, which we denote R_{off} . It is straightforward to show that in the linear regime in which we operate the ratio $R_{\text{on}}/R_{\text{off}}$ is equivalent to the ratio of the total cross-sections in each case, that is, $\sigma_{\text{on}}/\sigma_{\text{off}}$. We repeat this entire sequence >50 times and thus obtain a series of data points $R_{\text{on},1}$, $R_{\text{off},1}$, $R_{\text{on},2}$, $R_{\text{off},2}$, To eliminate monotonic, time-varying systematic errors, we compute and average the ratios $R_{\text{on},1}/R_{\text{off},1}$, $R_{\text{on},2}/R_{\text{off},2}$, ... as well as the ratios $R_{\text{on},2}/R_{\text{off},1}$, $R_{\text{on},3}/R_{\text{off},2}$, We then average the two results and repeat the experiment at various photodissociation wavelengths. These data are plotted as blue circles in Fig. 2b. Also plotted (black circles) are the results of a control analysis in which we instead perform the same calculation for the ratios $R_{\text{on},1}/R_{\text{on},2}$, $R_{\text{on},3}/R_{\text{on},4}$, ... as well as $R_{\text{off},1}/R_{\text{off},2}$, $R_{\text{off},3}/R_{\text{off},4}$, ... and average the results. Because these controls are comparisons of the same data type, they should yield results consistent with unity, and their variability about this value can be used as an indication of the variability in the actual data. Their close agreement with unity shows that the deviation from unity of the true data is due to MOT-induced vibrational relaxation and is not a statistical coincidence.

The number of Ba^+ ions created in each instance of photodissociation can be obtained using the following rate equations:

$$\begin{aligned} \frac{dN_i}{dt} &= \rho\Phi(k_{i+1,i}N_{i+1} - k_{i,i-1}N_i) - \frac{\sigma_i(\lambda)I}{h\nu}N_i \\ &\quad + \sum_{j>i} A_{ji}N_j - \sum_{j<i} A_{ij}N_i + \sum_{j\neq i} B_{ij}\rho(\omega)(N_j - N_i) \\ \frac{dN_{\text{Ba}}}{dt} &= \sum_i \frac{\sigma_i(\lambda)I}{h\nu}N_i \end{aligned} \quad (2)$$

Here N_i denotes the number of BaCl^+ molecules in the i th vibrational level (we have averaged over the rotational states) and N_{Ba} is the number of Ba^+ ions.

The first term in the expression for dN_i/dt (equation (2)) is due to the MOT-induced quenching: $k_{i+1,i}$ is the vibrational quenching rate constant for transition from the $(i+1)$ th level to the i th level, ρ is the peak density of the MOT and Φ is a factor that quantifies the degree of overlap of the atom and ion clouds. We define $\Phi = \int \bar{\rho}_{\text{Ca}}(r)\bar{\rho}_i(r) dr$, where $\bar{\rho}_{\text{Ca}}$ is the unit-peak normalized Ca atom density and $\bar{\rho}_i$ is the unit-integral normalized BaCl^+ density. Thus, $0 < \Phi < 1$. In the liquid phase in which we operate, Φ quantifies the effective MOT density experienced by each ion as they move in the trap.

The second term in the expression for dN_i/dt results from photodissociation (σ_i is the wavelength-dependent cross-section for the i th level, I is the PDL intensity and ν is the laser frequency). The last three terms are due to spontaneous emission and BBR-induced stimulated emission and absorption written in terms of the spectral energy density, $\rho(\omega)$, and the Einstein A and B coefficients³².

Calculating these coefficients requires knowledge of the dipole moment of the electronic ground state of BaCl^+ , $A_{ij} = (2/3)(\omega_{ij}^3\mu_{ij}^2/\epsilon_0\hbar c^3)$, where $\mu_{ij} = \langle j|d(R)|i\rangle$, $|n\rangle$ represents the rovibrational wavefunction of the n th rovibrational level of the ground electronic state, ω_{ij} is the angular frequency of the transition, ϵ_0 is the vacuum permittivity and c is the speed of light. We calculated this dipole moment function, which was previously unknown, using a non-relativistic, multiconfigurational, second-order perturbation theory (CASPT2) implemented in the MOLCAS software suite. The result is shown in Fig. 1b. With this in hand, the necessary rates are calculated according to ref. 32, and are ~ 1 Hz. Finally, we incorporate the lowest four vibrational levels into the model and ignore terms where $|i-j| > 1$. This model, along with the calculated effective density of the MOT²⁶ for each data point, allows a determination of the vibrational quenching rate constant, k , as well as the ground-state vibrational population, $P_{v=0}$.

32. Lefebvre-Brion, H. & Field, R. W. *The Spectra and Dynamics of Diatomic Molecules* Ch. 5 (Academic, 2004).

Deglacial pulses of deep-ocean silicate into the subtropical North Atlantic Ocean

A. N. Meckler¹, D. M. Sigman², K. A. Gibson³, R. François⁴, A. Martínez-García¹, S. L. Jaccard¹, U. Röhl⁵, L. C. Peterson³, R. Tiedemann⁶ & G. H. Haug¹

Growing evidence suggests that the low atmospheric CO₂ concentration of the ice ages resulted from enhanced storage of CO₂ in the ocean interior, largely as a result of changes in the Southern Ocean¹. Early in the most recent deglaciation, a reduction in North Atlantic overturning circulation seems to have driven CO₂ release from the Southern Ocean^{2–5}, but the mechanism connecting the North Atlantic and the Southern Ocean remains unclear. Biogenic opal export in the low-latitude ocean relies on silicate from the underlying thermocline, the concentration of which is affected by the circulation of the ocean interior. Here we report a record of biogenic opal export from a coastal upwelling system off the coast of northwest Africa that shows pronounced opal maxima during each glacial termination over the past 550,000 years. These opal peaks are consistent with a strong deglacial reduction in the formation of silicate-poor glacial North Atlantic intermediate water² (GNAIW). The loss of GNAIW allowed mixing with underlying silicate-rich deep water to increase the silicate supply to the surface ocean. An increase in westerly-wind-driven upwelling in the Southern Ocean in response to the North Atlantic change has been proposed to drive the deglacial rise in atmospheric CO₂ (refs 3, 4). However, such a circulation change would have accelerated the formation of Antarctic intermediate water and sub-Antarctic mode water, which today have as little silicate as North Atlantic Deep Water and would have thus maintained low silicate concentrations in the Atlantic thermocline. The deglacial opal maxima reported here suggest an alternative mechanism for the deglacial CO₂ release^{5,6}. Just as the reduction in GNAIW led to upward silicate transport, it should also have allowed the downward mixing of warm, low-density surface water to reach into the deep ocean. The resulting decrease in the density of the deep Atlantic relative to the Southern Ocean surface promoted Antarctic overturning, which released CO₂ to the atmosphere.

Palaeo-oceanographic data show that during Heinrich stadial 1 (HS1), ~17.5 kyr ago, there was an abrupt decrease in GNAIW formation², widely believed to be caused by a meltwater discharge to the high-latitude North Atlantic. At the same time, Antarctica and the Southern Ocean warmed and the atmospheric CO₂ concentration rose, factors that are central to the view that changes in the Northern Hemisphere triggered those in the Southern Hemisphere, leading to the release of CO₂ from the Southern Ocean. One proposed mechanism is that North Atlantic cooling during HS1 drove the Southern Hemisphere westerly winds southwards, increasing the upwelling of deep water, which released CO₂ (refs 3, 4). Increased wind-driven upwelling during deglaciations would also have increased the formation rates of sub-Antarctic mode water (SAMW) and Antarctic intermediate water (AAIW), which are the main subpolar sources for low-latitude thermocline waters. Indeed, a number of deglacial observations have been attributed to such a change (see, for example, ref. 7). However, other studies have found evidence against

accelerated SAMW or AAIW flow into the north and south Atlantic during HS1^{8–10}.

At present, silicate concentrations in SAMW and AAIW are relatively low. This has been explained as the result of a relatively deep dissolution profile for sinking biogenic silica, a high silicate-to-nitrate uptake ratio by diatoms, or both. These two factors reduce the silicate concentration in Antarctic and sub-Antarctic surface water that is then subducted in AAIW and SAMW, respectively¹¹. The paucity of silicate in the modern low-latitude thermocline has inspired the use of biogenic opal export as a measure of the input of silicate-rich deep waters into the overlying thermocline¹².

We take the analogous approach of using opal export to reconstruct past changes in the communication between the thermocline and the deeper ocean. A 550-kyr-long sediment record from the subtropical North Atlantic, at Ocean Drilling Program (ODP) Site 658, is analysed for changes in opal export. ODP Site 658 (20° 44' 57'' N, 18° 34' 51'' W, 2,263-m water depth) is situated in the Mauritanian upwelling cell off northwest Africa (Fig. 1). The X-ray fluorescence (XRF) scanning-derived element log-ratio of silica to aluminium counts (ln(Si/Al)) is used as a proxy for sediment opal content. For the last glacial termination, we also report new ²³⁰Th-normalized opal flux data from nearby core MD03-2705 (18° 05' 49'' N, 21° 09' 11'' W, 3,085-m water depth).

The ln(Si/Al) record is dominated by pronounced peaks during each glacial termination (Fig. 2b), reflecting maxima in preserved opal flux (Fig. 3e). Additional precession-related cyclicity is observed, which is probably due to variations in quartz content caused by climate-related changes in grain size (Supplementary Fig. 1). The deglacial peaks, however, correspond well to maxima in opal content in the same core¹³ (Fig. 2c). The opal peaks coincide with maxima in overall phytoplankton productivity as inferred from chlorin concentrations¹⁴, confirming that the sedimentary opal maxima mainly reflect increased production and not preservation (Supplementary Discussion 1). Chlorin concentration peaks are also observed at other times during glacial cycles but in those cases are not accompanied by increased opal concentrations. This fact, together with the opal peaks not being consistently associated with surface water cooling¹⁵ (Supplementary Fig. 2), suggests that the opal peaks do not reflect stronger local upwelling but rather increased silicate concentration in the upwelled water. Focusing on the most recent deglaciation (Fig. 3), we observe two maxima in ln(Si/Al) and ²³⁰Th-normalized opal flux during HS1 and the Younger Dryas stadial. Similar changes in opal production are also seen in other data from ODP Site 658, other cores nearby^{16,17} (Fig. 3) and other sites in the North Atlantic^{18,19}, indicating that the Site 658 opal maxima do not record solely local processes. In addition, each opal peak at Site 658 corresponds to a deglacial maximum in Southern Ocean export production as inferred from Ba/Fe ratios²⁰ (Fig. 2d).

We interpret the deglacial opal peaks as resulting from the mixing of silicate-rich abyssal waters into the low-latitude Atlantic thermocline. During the Last Glacial Maximum (LGM), Atlantic abyssal water (that

¹Geological Institute, ETH Zürich, 8092 Zürich, Switzerland. ²Department of Geosciences, Princeton University, New Jersey 08544, USA. ³Rosenstiel School of Marine and Atmospheric Science, University of Miami, Florida 33149, USA. ⁴Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ⁵MARUM – Center for Marine Environmental Sciences, University of Bremen, 28359 Bremen, Germany. ⁶Alfred Wegener Institute, 27568 Bremerhaven, Germany.

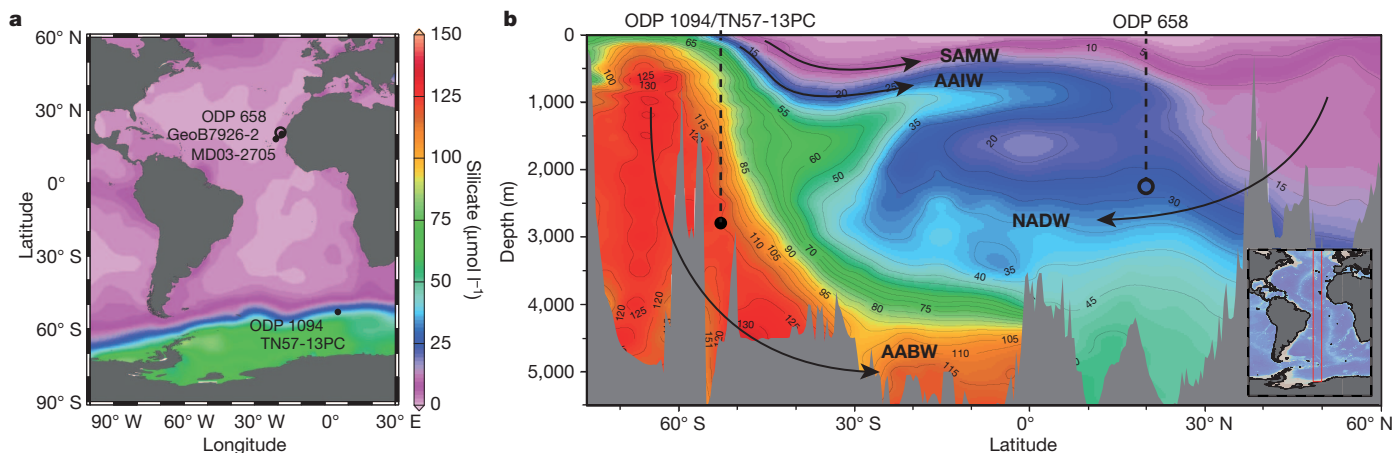


Figure 1 | Silicate concentration and locations of study sites. Open circle, ODP Site 658; filled circles: core MD03-2705 and other sites referred to in the text. **a**, Map of annual average surface silicate concentration²⁷, indicating the scarcity of this nutrient in most of the present-day Atlantic Ocean. **b**, Meridional depth section of silicate concentration in the central Atlantic

is, water below 2.5 km) apparently had a high silicate concentration, similar to that of modern Antarctic bottom water²¹ (AABW) (Fig. 4a). The strong reduction in GNAIW during HS1 allowed this silicate-rich water to shoal and mix upwards into the thermocline without the silicate being swept away by silicate-depleted intermediate-depth flows (Fig. 4b). We argue that this is a more plausible scenario than an increase in the silicate concentration or formation rate of SAMW or AAIW for the following reasons.

First, modern SAMW and AAIW, although rich in nitrate and phosphate, are much lower in silicate than is deep Southern Ocean water (~ 10 and $\sim 25 \mu\text{M}$ versus $125 \mu\text{M}$, respectively; Fig. 1). During glacial terminations, opal accumulation rates at ODP Site 658 increase up to tenfold, from glacial and interglacial background values of less than $0.2 \text{ g cm}^{-2} \text{ kyr}^{-1}$ to more than $2 \text{ g cm}^{-2} \text{ kyr}^{-1}$ during termination IV¹³ (Fig. 2). Even allowing for enhanced opal preservation under higher opal fluxes, this dynamic range would have required a

Ocean²⁷ (averaged over the area between the two red lines in the inset map), illustrating the silicate concentration in major present-day water masses (arrows) and the locations of Site 658 and Southern Ocean core TN57-13PC (same location as ODP Site 1094). The maps were made using the Ocean Data View software³⁰.

drastic rise in the preformed silicate concentration of SAMW and AAIW (Supplementary Discussion 2). The strong wind-driven overturning of the modern Southern Ocean fails to deliver substantial concentrations of silicate into SAMW and AAIW today, making such a rise unlikely. This view is only strengthened by the early decline in dust-borne iron during deglaciations, which may have caused a deglacial increase in the Si/N uptake ratio of Southern Ocean diatoms²², and by the early deglacial increase in Antarctic air and Southern Ocean temperature²³, which should have reduced the capacity of sea ice cover to limit silicate consumption by Antarctic diatoms.

Second, data on changes in the ocean interior favour our explanation over increased influence of a high-silicate form of AAIW or SAMW. The opal maxima consistently correspond to minima in the ^{13}C -to- ^{12}C ratio ($\delta^{13}\text{C}$) of benthic foraminifera at Site 658 (Supplementary Fig. 2),

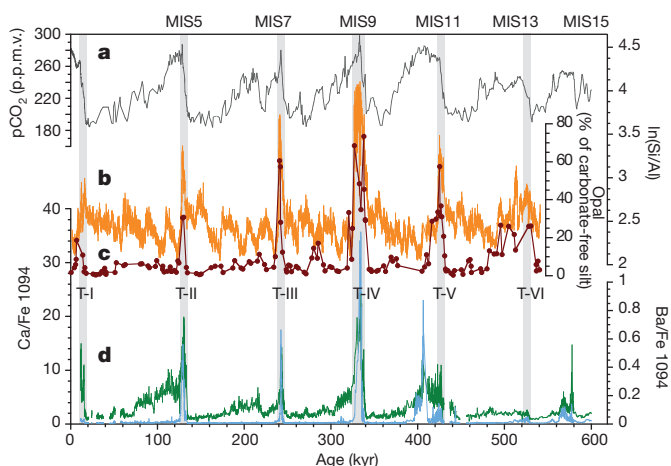


Figure 2 | Opal record from ODP Site 658 compared with atmospheric CO_2 concentration and productivity and carbonate preservation in the Southern Ocean. **a**, pCO_2 record from EPICA Dome C³¹; **b**, $\ln(\text{Si}/\text{Al})$ record from ODP Site 658, off Mauritania; **c**, percentage opal record from Site 658 (ref. 13); **d**, Ba/Fe (green), a proxy for marine export production, and Ca/Fe (blue), a proxy for carbonate preservation, from Southern Ocean ODP Site 1094 (ref. 20). The deglacial peaks in export production at Site 1094 are interpreted as increased nutrient content in upwelled waters, and the carbonate preservation peaks are taken to reflect CO_2 release from the deep ocean. Grey shading indicates glacial terminations I to VI, and MIS 5 to MIS 15 are interglacial periods.

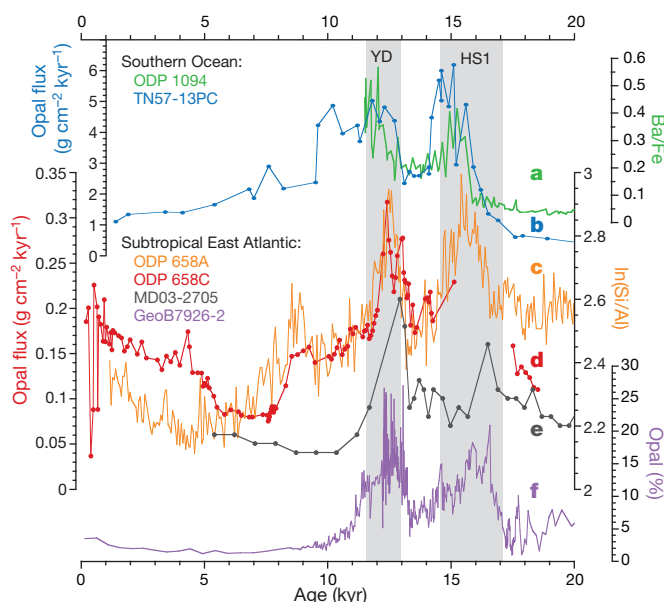


Figure 3 | Comparison of opal records for the last glacial termination. **a**, ^{230}Th -normalized opal flux from core TN57-13PC (ref. 4); **b**, Ba/Fe record from ODP Site 1094 (ref. 20); **c**, $\ln(\text{Si}/\text{Al})$ record from ODP Site 658 (Hole A; this study); **d**, ^{230}Th -normalized opal flux from Site 658 (Hole C; ref. 16); **e**, ^{230}Th -normalized opal flux from core MD03-2705 (this study); **f**, percentage opal in core GeoB7926-2 (ref. 17). The core locations are shown in Fig. 1a. YD, Younger Dryas stadial.

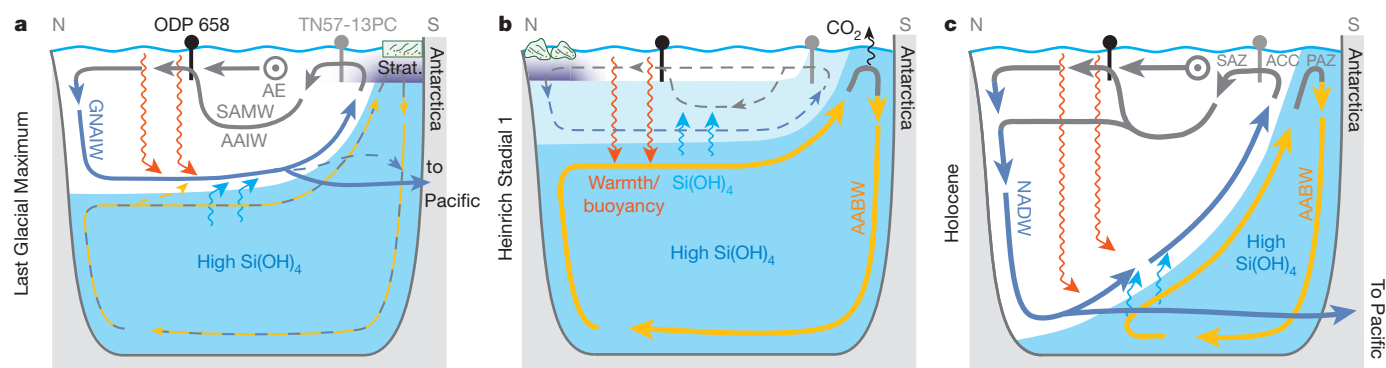


Figure 4 | Schematic north-south section of the Atlantic Ocean at different stages of the most recent glacial termination. **a**, The Last Glacial Maximum; **b**, Heinrich Stadial 1; **c**, the Holocene (current interglacial period). AE, Agulhas eddies (inflow of surface water from the Indo-Pacific into the Atlantic); ACC, Antarctic circumpolar current; PAZ, polar Antarctic zone; SAZ, sub-Antarctic zone. Line colours of interior flows indicate their ventilation source region: blue, NADW or GNAIW; yellow, AABW. Line thickness changes among panels denote changes in flow rate. Blue shading in the interior reflects the concentration of dissolved silicate ($\text{Si}(\text{OH})_4$). The black and grey dots indicate the locations of ODP Site 658 (this study) and core TN57-13PC (ref. 4), respectively, with connected lines representing the source of the upwelled water. Purple shading indicates increased polar water column stratification. In the LGM state (**a**), the replacement of NADW by GNAIW was accompanied by

shoaling of the upper boundary of silicate-rich southern-sourced deep water. During early deglaciation (HS1; **b**), a reduction in GNAIW formation in the North Atlantic allowed the upward mixing of abyssal water rich in silicate (blue wavy arrows) to increase the silicate concentration of intermediate and thermocline waters (light blue shading), resulting in the observed peaks in opal production at Site 658 and in TN57-13PC. The simultaneous downward mixing of warmth and, thus, buoyancy (red wavy arrows) reduced the density of deep water, stimulating increased formation of AABW and leading to CO_2 release through the PAZ. The interglacial (Holocene) state (**c**) is characterized by strong deep and intermediate water flows from both north and south, which transport away the silicate mixed up from southern-sourced abyssal water, preventing it from reaching the low-latitude thermocline.

probably reflecting an increased dominance of nutrient-rich water from the deeper Atlantic. During HS1 in both the north and south Atlantic, the strongest part of the vertical $\delta^{13}\text{C}$ gradient shoaled to ~ 1 km, with no sign of a focused mid-depth $\delta^{13}\text{C}$ anomaly entering from the south at the depths of AAIW²⁴, arguing against an increase in the rate of AAIW or SAMW inflow to the Atlantic during HS1. Consistent with this finding, Cd/Ca data from a South Atlantic site at the depth of AAIW show no discernable increase from low glacial Cd/Ca values until after HS1¹⁰. Studies of Cd/Ca and Nd isotopes in the Florida straits at the depth of modern SAMW and AAIW (~ 750 m) suggest that the flow of these southern-sourced water masses into the North Atlantic was low during the last ice age and did not increase until after HS1, during the Bølling–Allerød warm period, with a clear interruption during the Younger Dryas^{8,9}. In fact, the Nd isotope data (and less strongly, the Cd/Ca data) suggest a decline in AAIW penetration during HS1 relative to LGM conditions. Because modern SAMW and AAIW are incorporated into North Atlantic Deep Water (NADW; Fig. 4c), the apparent restart of NADW during the Bølling–Allerød warm period has been proposed as the cause for increased SAMW or AAIW flow into the North Atlantic at that time⁸. We note that a number of studies interpret mid-depth changes during HS1 (in Cd/Ca, Si isotopes, Nd isotopes and water-mass radiocarbon content) as resulting from increased AAIW formation, but these data are equally well, or better, explained by increased upward incursion of the underlying southern-sourced abyssal water (Supplementary Discussion 2).

Third, in the Pacific Ocean, deglacial opal flux maxima have been observed during some but not all of the past five glacial terminations²⁵. According to our proposed mechanism and as observed during the most recent deglaciation, the deglacial reduction in GNAIW should have affected intermediate-depth circulation in the Atlantic more strongly than in the Pacific. In contrast, if the opal peaks were the result of silicate-rich deglacial forms of SAMW and AAIW, they should have been as robust in the Pacific as in the Atlantic.

One observation that our proposed deglacial mechanism does not explain is the continued deposition of opal at ODP Site 658 through the early interglacial period of Marine Isotope Stage 9 (MIS 9; Fig. 2b). NADW formation was probably active during MIS 9, which under normal circumstances would have isolated the thermocline from the high silicate concentration of southern-sourced deep waters. An intriguing possibility for the cause of the high interglacial opal flux of MIS 9

is that NADW flowed below, rather than above, all southern-sourced waters. That is, AABW and AAIW may have been merged as a single mid-depth water mass during MIS 9, with NADW below it. MIS 9 had the highest pCO_2 levels of recent interglacials, which may have rendered the high latitudes unusually warm relative to the global mean. These conditions would have been especially effective at reducing the density of AABW, because its formation is probably more sensitive to polar temperatures than is that of NADW, which forms at lower latitude and is also more strongly tied to salinity.

The reconstructed deglacial opal maxima in the subtropical North Atlantic clarify multiple aspects of the mechanism of deglaciation. Because GNAIW formation and export must have decreased drastically to produce these pronounced peaks in opal flux, their systematic occurrence indicates that the early reduction in GNAIW formation is a universal aspect of deglaciations and is very probably central to the sequence of events underlying glacial terminations. Furthermore, the evidence against an increase in SAMW or AAIW formation makes our observations inconsistent with the ‘wind-shift’ hypothesis for the early deglacial CO_2 release^{3,4}. The increased Antarctic Ekman upwelling that has been proposed to drive the CO_2 release is physically coupled to the equatorward transport of Antarctic surface waters and ‘Ekman pumping’ farther north²⁶. An increased northward flow would have gone into the formation of additional SAMW or AAIW, or some shallower, even more silicate-deplete, subsurface water mass. Any of these contributions to the upper water column of the low-latitude Atlantic would have increased the vertical separation between the low-latitude surface and the silicate-rich deep water, counteracting the upward penetration of silicate into the thermocline that is required to generate the deglacial opal peaks in the low-latitude Atlantic. In addition, stronger Southern Hemisphere westerly winds during HS1 would have steepened the isopycnals in the subpolar Southern Hemisphere, thickening the Atlantic thermocline²⁶. This again would have worked to separate the shallow tropical Atlantic from silicate-rich deep water, counteracting the mechanisms required to generate the observed opal maxima.

In Antarctic sediments, the higher interglacial concentrations of opal and Ba (as in the Ba/Fe record of ODP Site 1094; Fig. 2d) compared with glacial values are interpreted as evidence of increased Antarctic overturning¹. As a corollary, the deglacial opal maxima observed in core TN57-13PC (cored at the same location as Site 1094; Fig. 3a) have been taken as indicating transient deglacial accelerations in Antarctic

overturning that then subsided to modestly increased overturning during the full interglacials⁴. However, the pronounced deglacial opal maxima of TN57-13PC very probably share the same cause as those at ODP Site 658: a deglacial increase in the silicate concentration of Atlantic mid-depth waters. At the site of TN57-13PC, the water at ~200-m depth, which is a major source of the nutrients imported into the euphotic zone, is on a density surface that today sits in the upper half of NADW in the low-latitude Atlantic (~1,800 m; ref. 27). The reduction in GNAIW formation during HS1 and the Younger Dryas, by allowing AABW to shoal, would have increased the silicate concentration on this density surface throughout the Atlantic. In this light, the persistent occurrence of deglacial export production maxima at Site 1094 (Fig. 2d) provides a further indication of the systematic loss of North Atlantic ventilation at glacial terminations.

Lacking the westerly-wind-driven mechanism, we are left with the question of how the deglacial reduction in GNAIW led to an increase in Antarctic overturning and CO₂ release. The deglacial opal maxima provide support for a mechanism that was first raised in the context of the Younger Dryas⁶: just as the loss of GNAIW would have allowed the upward transport of silicate from the abyssal Atlantic to the low-latitude thermocline, it also would have permitted the downward mixing of warm surface water to reach the southern-sourced deep waters (Fig. 4b, red wavy arrows). The resulting deep 'density vacuum'⁶, that is, a reduction in the density of deep water that stimulates its replacement by new, denser water, would have encouraged deep-water formation in the Antarctic, releasing CO₂ to the atmosphere⁵ (Fig. 4b). Early deglacial warming in the ocean interior has been documented²⁸. Although this warming is plausibly interpreted to indicate an Antarctic trigger for deglaciation²⁸, most records suggest that Antarctic air temperature increased later^{23,29}. Thus, an alternative view is that the HS1 shutdown in North Atlantic ventilation led to the ocean interior warming, which, in turn, triggered Antarctic overturning, causing Antarctica to warm and the atmospheric CO₂ concentration to rise.

METHODS SUMMARY

Elemental ratios at ODP Site 658 were determined by XRF scanning of the split core surface using an Avaatech core scanner. Measurements were taken continuously at a resolution of 1 cm. Cores from all three parallel holes of Site 658 were used and a new composite depth was derived from the XRF data. The age model for Site 658 was constructed on the basis of oxygen isotope stratigraphy and additional matching of scanning XRF data to published records from Site 658 and a nearby core (Supplementary Fig. 3). For core MD03-2705, the published age model was used. We derived ²³⁰Th-normalized opal flux data for core MD03-2705 from biogenic opal concentrations and excess ²³⁰Th activities, determined by isotope dilution and measurement on a single-collector inductively-coupled-plasma sector-field mass spectrometer.

Full Methods and any associated references are available in the online version of the paper.

Received 5 September 2012; accepted 12 February 2013.

1. Sigman, D. M., Hain, M. P. & Haug, G. H. The polar ocean and glacial cycles in atmospheric CO₂ concentration. *Nature* **466**, 47–55 (2010).
2. McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
3. Toggweiler, J. R. & Lea, D. W. Temperature differences between the hemispheres and ice age climate variability. *Paleoceanography* **25**, PA2212 (2010).
4. Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO₂. *Science* **323**, 1443–1448 (2009).
5. Sigman, D. M., de Boer, A. M. & Haug, G. H. in *Ocean Circulation: Mechanisms and Impacts* (eds Schmittner, A., Chiang, J. C. H. & Hemming, S. R.) 335–349 (Geophys. Monogr. Ser. 173, Am. Geophys. Union, 2007).
6. Broecker, W. S. Paleocene circulation during the last deglaciation: a bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
7. Rickaby, R. E. M. & Elderfield, H. Evidence from the high-latitude North Atlantic for variations in Antarctic Intermediate water flow during the last deglaciation. *Geochim. Geophys. Res.* **6**, Q05001 (2005).
8. Came, R. E., Oppo, D. W., Curry, W. B. & Lynch-Stieglitz, J. Deglacial variability in the surface return flow of the Atlantic meridional overturning circulation. *Paleoceanography* **23**, PA1217 (2008).

9. Xie, R. C., Marcantonio, F. & Schmidt, M. W. Deglacial variability of Antarctic Intermediate Water penetration into the North Atlantic from authigenic neodymium isotope ratios. *Paleoceanography* **27**, PA3221 (2012).
10. Came, R. E., Oppo, D. W. & Curry, W. B. Atlantic Ocean circulation during the Younger Dryas: insights from a new Cd/Ca record from the western subtropical South Atlantic. *Paleoceanography* **18**, 1086 (2003).
11. Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2004).
12. Gnanadesikan, A. & Toggweiler, J. R. Constraints placed by silicon cycling on vertical exchange in general circulation models. *Geophys. Res. Lett.* **26**, 1865–1868 (1999).
13. Tiedemann, R., Sarnthein, M. & Stein, R. in *Proc. ODP Sci. Res. Vol. 108* (eds Ruddiman, W. F. & Sarnthein, M.) 241–277 (Texas A & M Univ., 1989).
14. Harris, P. G. *et al.* Chlorin accumulation rate as a proxy for Quaternary marine primary productivity. *Nature* **383**, 63–65 (1996).
15. Eglinton, G. *et al.* Molecular record of secular sea-surface temperature-changes on 100-year timescales for glacial terminations I, II and IV. *Nature* **356**, 423–426 (1992).
16. Adkins, J., deMenocal, P. & Eshel, G. The "African humid period" and the record of marine upwelling from excess Th-230 in Ocean Drilling Program Hole 658C. *Paleoceanography* **21**, PA4203 (2006).
17. Romero, O. E., Kim, J. H. & Donner, B. Submillennial-to-millennial variability of diatom production off Mauritania, NW Africa, during the last glacial cycle. *Paleoceanography* **23**, PA3218 (2008).
18. Bradtmiller, L. I., Anderson, R. F., Fleisher, M. Q. & Burckle, L. H. Opal burial in the equatorial Atlantic Ocean over the last 30 ka: implications for glacial-interglacial changes in the ocean silicon cycle. *Paleoceanography* **22**, PA4216 (2007).
19. Gil, I. M., Keigwin, L. D. & Abrantes, F. G. Deglacial diatom productivity and surface ocean properties over the Bermuda Rise, northeast Sargasso Sea. *Paleoceanography* **24**, PA4101 (2009).
20. Jaccard, S. L. *et al.* The distinct roles of the Antarctic and Subantarctic Zones in ocean productivity and atmospheric CO₂ over the past million years. *Science*. (in the press).
21. Marchitto, T. M., Oppo, D. W. & Curry, W. B. Paired benthic foraminiferal Cd/Ca and Zn/Ca evidence for a greatly increased presence of Southern Ocean Water in the glacial North Atlantic. *Paleoceanography* **17**, 1038 (2002).
22. Brzezinski, M. A. *et al.* A switch from Si(OH)₄ to NO₃[−] depletion in the glacial Southern Ocean. *Geophys. Res. Lett.* **29**, 5 (2002).
23. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793–796 (2007).
24. Oppo, D. W. & Curry, W. B. Deep Atlantic circulation during the Last Glacial Maximum and deglaciation. *Nature Edu. Knowl.* **3**, 1 (2012).
25. Hayes, C. T., Anderson, R. F. & Fleisher, M. Q. Opal accumulation rates in the equatorial Pacific and mechanisms of deglaciation. *Paleoceanography* **26**, PA1207 (2011).
26. Marshall, J. & Speer, K. Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geosci.* **5**, 171–180 (2012).
27. Levitus, S. in *NOAA Atlas NESDIS 68* 184 (US Govt Printing Office, 2009).
28. Stott, L., Timmermann, A. & Thunell, R. Southern hemisphere and deep-sea warming led deglacial atmospheric CO₂ rise and tropical warming. *Science* **318**, 435–438 (2007).
29. Shakun, J. D. *et al.* Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484**, 49–54 (2012).
30. Schlitzer, R. Electronic atlas of WOCE hydrographic and tracer data now available. *Eos Trans. AGU* **81**, 45 (2000).
31. Siegenthaler, U. *et al.* Stable carbon cycle-climate relationship during the late Pleistocene. *Science* **310**, 1313–1317 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research used samples provided by the ODP, which is sponsored by the US NSF and participating countries under the management of the Joint Oceanographic Institutions. XRF data were acquired at the XRF Core Scanner Lab at MARUM – Center for Marine Environmental Sciences, University of Bremen, with support from the DFG-Leibniz Center for Surface Process and Climate Studies at the University of Potsdam. Further support was provided by the US NSF through grant OCE-1060947 to D.M.S. and by NSERC and CFCAS to R.F. We thank V. Lukies for XRF scanning technical support and T. Westerhold for help in deriving the new composite depth. M. Soon and T. Kane are acknowledged for opal and excess ²³⁰Th measurements, respectively, in core MD03-2705.

Author Contributions A.N.M. and G.H.H. designed the study; A.N.M. collected the XRF data at ODP Site 658; U.R. facilitated and oversaw the XRF scanning; R.F. contributed the ²³⁰Th-normalized opal flux data; R.T. provided data and background knowledge on Site 658; A.N.M., S.L.J. and A.M.G. undertook the comparisons with Southern Ocean data; A.N.M., D.M.S. and G.H.H. wrote the first draft of the manuscript; and K.A.G. and L.C.P. provided input regarding tropical Atlantic observations. All authors contributed to the interpretation and the preparation of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.N.M. (nele.meckler@erdw.ethz.ch).

METHODS

XRF scanning (ODP Site 658). The scanning XRF measurements at ODP Site 658 were performed with an Avaatech XRF Core Scanner at MARUM, University of Bremen. Data were obtained at a resolution of 1 cm over an area of 1.2 cm² directly at the split core surface of the archive half, with different settings for light elements such as Al and Si (10 kV, 20 s, and 150 mA) and heavy elements such as Zr (50 kV, 20 s and 80 mA). The core surface was covered with 4-µm-thick SPEXCerti Prep Ultralene1 foil to avoid contamination of the XRF measurement unit. The core scanner includes a Canberra X-PIPS Silicon Drift Detector (SDD; model SXD 15C-150-500) with 150-eV X-ray resolution, a Canberra Digital Spectrum Analyser DAS 1000 and an Oxford Instruments 100W Neptune X-ray tube with rhodium (Rh) target material. Raw data spectra were processed using the WIN AXIL (analysis of X-ray spectra by iterative least squares) package from Canberra Eurisys. Reported element count ratios were calculated from XRF signal intensities.

To bridge coring gaps and hiatuses, cores from all three parallel holes (A, B, and C) were used (Supplementary Fig. 2) and a new composite depth was derived from the XRF data. In cores from Hole C, the water content of the sediment is higher than in cores from the other two holes because the Hole C cores were frozen after drilling to preserve organic molecules better and re-thawed much later in the core repository, leaving them less time to dry out. Because water content can affect the XRF measurements of light elements such as Al, we used this hole only where material from the other two holes was unavailable, and adjusted the Zr/Al and Si/Al ratios from Hole 658C by regression to Zr/Al and Si/Al ratios from Hole 658B for a section with overlapping data.

Grain size correction. Grain size at ODP Site 658 and in nearby cores has been shown to vary owing to changes in local climate^{13,32}. Minerals such as quartz (SiO₂) and zircon (ZrSiO₄) are resistant to weathering and are hence mainly associated with coarser grain sizes. Grain size therefore affects the abundance of Si and zirconium (Zr) relative to Al, which is mainly associated with clay minerals. At Site 658, ln(Zr/Al) exhibits very similar variability to ln(Si/Al) (Supplementary Fig. 1), except for the pronounced deglacial peaks. These variations also correspond well to the published low-resolution grain size record from the same core¹³ (Supplementary Fig. 1). The precession-related variability in ln(Si/Al) therefore probably reflects the abundance of quartz, not opal. When subtracting the two normalized XRF-derived time series, the grain-size-corrected (that is, quartz-corrected) ln(Si/Al) record agrees well with the published low-resolution opal record¹³ (Supplementary Fig. 1).

²³⁰Th normalized opal flux (core MD03-2705). Normalized opal fluxes were calculated by multiplying the opal fraction of the sediment by the ratio of expected

flux of ²³⁰Th scavenged from the water column to observed initial excess ²³⁰Th activities. Biogenic opal was determined by alkaline extraction of silica³³ and measurement of silica concentrations in the extract by molybdate-blue spectrophotometry. Replicate analyses indicate a reproducibility of ±3%. ²³⁰Th was measured by isotope dilution on a Finnigan 144 MAT Element 2 single-collector inductively-coupled-plasma sector-field mass spectrometer. The sediment samples (~0.3 g) were ground and spiked with ²²⁹Th before total dissolution in HNO₃, HF and HClO₄. An aliquot of the resulting solution was spiked with ²³⁶U and ²²⁹Th to measure ²³⁸U and ²³²Th. The remaining solution was used for ²³⁰Th separation by ion-exchange chromatography. Initial excess ²³⁰Th activities were obtained by correcting for the supported detrital ²³⁰Th assuming a ²³⁸U/²³²Th ratio of 0.7, ingrowth from authigenic uranium and radioactive decay since deposition. The expected excess ²³⁰Th activity was based on its production rate from ²³⁴U in the water column, given the water depth of the coring site.

Age models. The age model for the ODP Site 658 record was derived by transferring the published benthic δ¹⁸O record³⁴ onto the new composite depth and tuning to the same data on the age scale used in the LR04 benthic stack³⁵ (Supplementary Fig. 3). For the past 120 kyr, the age model was additionally refined by matching XRF Ca counts from Hole 658A and Hole 658B to percentage CaCO₃ data from Hole 658C, which has been radiocarbon-dated (ref. 36) (0–15 kyr ago), and by matching Zr/Al ratios to the grain-size-derived humidity index from neighbouring core GeoB7920-2 (ref. 32) (30–120 kyr ago). All data matching was performed using the software AnalySeries³⁷. For core MD03-2705, the published age model³⁸ was used.

32. Tjallingii, R. *et al.* Coherent high- and low-latitude control of the northwest African hydrological balance. *Nature Geosci.* **1**, 670–675 (2008).
33. Mortlock, R. A. & Froelich, P. N. A simple method for the rapid-determination of biogenic opal in pelagic marine-sediments. *Deep-Sea Res.* **36**, 1415–1426 (1989).
34. Sarnthein, M. & Tiedemann, R. Younger Dryas-style cooling events at glacial terminations I–VI at ODP site 658: associated benthic delta 13C anomalies constrain meltwater hypothesis. *Paleoceanography* **5**, 1041–1055 (1990).
35. Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic delta O-18 records. *Paleoceanography* **20**, PA1003 (2005).
36. deMenocal, P. *et al.* Abrupt onset and termination of the African Humid Period: rapid climate responses to gradual insolation forcing. *Quat. Sci. Rev.* **19**, 347–361 (2000).
37. Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *Eos Trans. AGU* **77**, 379 (1996).
38. Jullien, E. *et al.* Low-latitude “dusty events” vs. high-latitude “icy Heinrich events”. *Quat. Res.* **68**, 379–386 (2007).

Electrical image of passive mantle upwelling beneath the northern East Pacific Rise

Kerry Key¹, Steven Constable¹, Lijun Liu² & Anne Pommier³

Melt generated by mantle upwelling is fundamental to the production of new oceanic crust at mid-ocean ridges, yet the forces controlling this process are debated^{1,2}. Passive-flow models predict symmetric upwelling due to viscous drag from the diverging tectonic plates, but have been challenged by geophysical observations of asymmetric upwelling^{3–5} that suggest anomalous mantle pressure and temperature gradients^{2,6,7}, and by observations of concentrated upwelling centres⁸ consistent with active models where buoyancy forces give rise to focused convective flow². Here we use sea-floor magnetotelluric soundings at the fast-spreading northern East Pacific Rise to image mantle electrical structure to a depth of about 160 kilometres. Our data reveal a symmetric, high-conductivity region at depths of 20–90 kilometres that is consistent with partial melting of passively upwelling mantle^{9–11}. The triangular region of conductive partial melt matches passive-flow predictions, suggesting that melt focusing to the ridge occurs in the porous melting region rather than along the shallower base of the thermal lithosphere. A deeper conductor observed east of the ridge at a depth of more than 100 kilometres is explained by asymmetric upwelling due to viscous coupling across two nearby transform faults. Significant electrical anisotropy occurs only in the shallowest mantle east of the ridge axis, where high vertical conductivity at depths of 10–20 kilometres indicates localized porous conduits. This suggests that a coincident seismic-velocity anomaly¹² is evidence of shallow magma transport channels^{13,14} rather than deeper off-axis upwelling. We interpret the mantle electrical structure as evidence that plate-driven passive upwelling dominates this ridge segment, with dynamic forces being negligible.

We collected magnetotelluric data on the northern East Pacific Rise (EPR) to study the electrical structure of mantle upwelling, melt generation and magma transport to the crust at a mid-ocean ridge (Fig. 1). Magnetotelluric data record naturally occurring low-frequency electromagnetic variations that sense electrical conductivity, a physical property that primarily depends on temperature, mineralogy and the presence of fluids and their interconnectedness^{10,11}. The upwelling region beneath mid-ocean ridges is a particularly suitable target for magnetotelluric exploration, because small fractions of conductive partial melt can greatly increase bulk conductivity from the generally low conductivity of unmelted peridotite^{5,10,11,15}.

The fast spreading EPR separates the Pacific and Cocos plates, which diverge symmetrically to produce new sea floor at a half-spreading rate of 55 km Myr⁻¹, while the entire ridge system migrates to the north-west at 50 km Myr⁻¹ relative to a hotspot reference frame¹⁶ (Fig. 1). The 8–11° N EPR ridge segment became a locus of scientific investigation after the first direct observations of extensive sea-floor volcanism¹⁷. Vigorous hydrothermal venting¹⁸ and a continuous crustal magma chamber¹⁹ suggest this ridge segment taps a robust mantle melt supply, a notion reinforced by a sea-floor eruption that buried several actively recording seismometers²⁰. Seismic tomography shows the shallowest mantle at 9 km depth contains a low P-wave-velocity zone that follows the ridge crest along most of this segment, but which

deviates 5–10 km east of the ridge near our magnetotelluric profile (Fig. 1), suggesting that melt delivery to the base of the crust is locally focused eastward here¹². Despite a multitude of studies at this ridge segment, the deeper mantle upwelling region remained unexplored by a large geophysical array until this magnetotelluric study.

Our survey relied on broadband magnetotelluric instrumentation that is sensitive to the high frequencies required to constrain the shallow mantle¹⁵ and that requires only 10–20 days of recording time, making array surveys possible with a single research cruise. The 200-km profile of 29 magnetotelluric stations contains a broad aperture suitable for imaging mantle upwelling but also has closely spaced stations near the ridge axis to constrain melt focusing to the base of the crust (Fig. 1). For comparison, the earlier MELT experiment obtained sensitivity to deeper upwelling structure using year-long deployments of 19 long-period instruments along a more sparsely sampled 400-km transect⁵. We used regularized nonlinear data inversion to produce a smooth image of two-dimensional electrical resistivity constrained to about 160 km depth and to 100 km distance from the ridge axis (Fig. 2).

The inversion model reveals a prominent conductive region (<100 Ω m resistivity) centred beneath the ridge axis. Conductivity (the reciprocal of resistivity) is greatest at depths of about 20–90 km and is largely symmetric about the ridge axis. The top of the conductive region displays a broad triangular shape that deepens away from the ridge, reaching depths of about 80–100 km at 100 km distance. Above the sides of this conductor, the mantle becomes significantly more resistive (100–2,000 Ω m). The deeper mantle, at depths >100 km, is less resistive than 60 Ω m, but is asymmetric with resistivity as low as 10 Ω m to the east. Model perturbation tests show that the central and deeper conductive regions are required to obtain a good fit to the data (see Supplementary Figs 7 and 8), although the deeper conductor is less well constrained owing to its location near the edge of the magnetotelluric array and its depth. Anisotropy is insignificant throughout most of the model, except beneath the ridge at 10–20 km depth and 0–10 km position, where conductivity is greatest in the vertical direction (Supplementary Fig. 5).

These conductive regions indicate that the upwelling mantle is undergoing significant partial melting. Mineral physics studies find that upwelling will trigger decompression melting of dry peridotite at a depth of about 65 km, where the ridge adiabat crosses the peridotite solidus⁹. However, the deeper extent of our conductive region agrees with experimental studies of mid-ocean-ridge basalt samples that show the upper-mantle source region is slightly damp, with 50–200 parts per million (p.p.m.) H₂O (refs 21, 22), resulting in a deeper peridotite solidus that promotes melting to around 100 km depth⁹, depending on the amount of H₂O present. The presence of dissolved CO₂ has also been reported in mid-ocean-ridge contexts, and is thought to stabilize carbonated silicate partial melts at even greater depths^{23,24}, and the combination of CO₂ and H₂O may further enhance deep silicate melting²⁵.

Recent laboratory electrical conductivity studies have shown that small amounts of dissolved H₂O and CO₂ will increase the conductivity

¹Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA. ²Department of Geology, College of Liberal Arts and Sciences, University of Illinois, Urbana, Illinois 61801, USA. ³School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA.

of mantle partial melt^{10,11}. Because these volatiles are preferentially partitioned into melt, the earliest deep melts will be the most conductive, whereas later shallower melts will have lower volatile fractions and hence display a conductivity that is transitional between the volatile-influenced and dry endmembers¹¹. The low resistivity of 1–10 Ω m in the main conductive body is consistent with up to a few volume per cent partial melt, but the highly conductive core of the anomaly (0.3 Ω m) requires a significantly higher melt fraction (>10 vol.%)^{10,11}. However, we warn against over-interpreting this high conductivity, because magnetotelluric data are primarily sensitive to the integrated conductance of conductive bodies; the inversion may be trading-off a narrow size for this core region with a compensating higher conductivity.

Because carbonate melts are much more conductive than silicate melt¹¹, the 10–60 Ω m resistivity beneath 100 km depth can be explained by a much lower amount of carbonate partial melt (<0.5 vol.%) that is stable at this depth. The moderately resistive gap separating the two conductive regions at around 100 km depth is intriguing, because the deeper carbonate melts should be migrating upward and enhancing conductivity there. However, experimental results have shown that when both silicate and carbonate melts are present, the solid matrix is preferentially wetted by silicate melt, and therefore carbonate would not necessarily be well connected²⁶; it may also be diluted by the less conductive silicate melt¹¹. We suggest that this gap represents the depth where silicate melt begins to dominate over carbonate melt, leading to an overall lower conductivity.

The melting region at depths shallower than 100 km in our study is unusually symmetric compared to geophysical observations at other mid-ocean ridges^{1,4,8}, but is juxtaposed with deeper asymmetric melting below 100 km. Although viscous drag from symmetrically diverging tectonic plates is thought to be the main driving force in mantle upwelling, asymmetric observations have required additional mechanisms in order to generate melt unevenly across the ridge. A strong

westward asymmetry observed at the southern EPR was conjectured to arise from mantle pressure or temperature gradients^{6,7}. Localized centres of upwelling observed in the Gulf of California do not align with the heavily segmented ridge axis, suggesting that melt may be dynamically concentrated through buoyant convective flow⁸.

We carried out a three-dimensional flow simulation to study whether the deep asymmetry in our model could be explained solely by passive flow, or requires the addition of dynamic forces or lateral temperature and pressure gradients. Absolute plate motions were imposed as a boundary condition on a simulation domain that included both the Clipperton and Siqueiros transform offsets. Figure 3 shows the vertical velocity component of flow induced by the plate motions. In the upwelling mantle, melting occurs in proportion to the vertical velocity due to decompression but also depends on composition and adiabatic temperature. Because the melting rate for hydrous, carbonated peridotite is poorly known, we use vertical velocity as a proxy indicator for where deep incipient melts could be generated.

The deep upwelling shows an asymmetry between the transform offsets, with vertical velocities of 0.5–2.5 cm yr⁻¹ biased to the east of the ridge axis at 100–250 km depth. Although we included westward ridge migration in the simulation via the absolute plate motions, it does not generate a significant asymmetry in melt production because migration only superimposes a horizontal shear to the deep flow^{6,7}. This is further supported by the lack of asymmetry in the simulation far north of the Clipperton transform and far south of the Siqueiros transform. Rather, the deep asymmetry arises from viscous coupling across the transform offsets, which is intensified because both offsets jump to the east and are in relatively close proximity. This upwelling asymmetry will generate more melt to the east, offering an explanation for our observed asymmetry. However, our model contains deep carbonate melts extending much further off-axis than the peak in upwelling. Because the magnetotelluric inversion is imaging the amount of partial melt present rather than the melting rate implied by the vertical

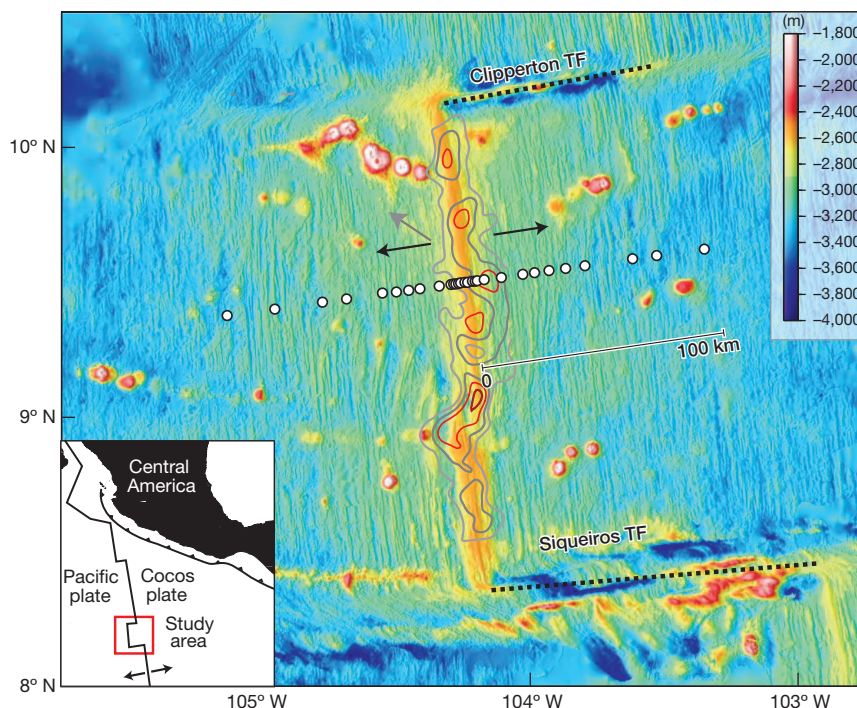


Figure 1 | Location of the magnetotelluric survey across the fast spreading East Pacific Rise. Twenty-nine sea-floor magnetotelluric stations (white circles) were deployed across the ridge axis at 9° 30' N, about 1,000 km southwest of Central America (inset; study area boxed in red). The Pacific and Cocos plates diverge symmetrically (black arrows) while the entire ridge system migrates to the northwest relative to a fixed hotspot reference frame¹⁶ (grey arrow). The Clipperton and Siqueiros transform faults (TF) bound this ridge

segment to the north and south. Slow seismic P-wave velocity contours (velocities <7.6 km s⁻¹) found by seismic tomography of the uppermost few kilometres of mantle¹² follow the ridge crest along most of the segment, but deviate to the east near the magnetotelluric profile. The colour scale shows sea-floor topography and the 100-km scale bar indicates the half-aperture of the magnetotelluric array.

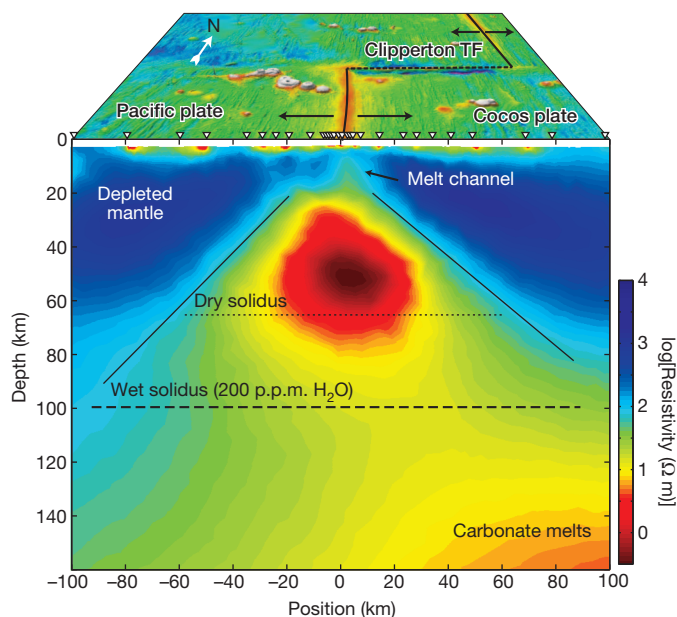


Figure 2 | Magnetotelluric resistivity image of mantle upwelling beneath the East Pacific Rise. Top, surface view of sea-floor topography. Main panel, colours show $\log[\text{resistivity } (\Omega \text{ m})]$ in the vertical direction obtained from nonlinear inversion of data from the sea-floor magnetotelluric stations (inverted triangles in surface view). Green to red colours indicate high conductivity (low resistivity) due to partial melts generated in the upwelling mantle. The full anisotropic resistivity results are shown in Supplementary Fig. 5. Solidus depths for dry and wet peridotite⁹ are shown as horizontal lines.

velocity, it is possible that this conductor represents melt that has ascended from greater depths; if deep off-axis melt is partially entrained in the mantle flow, it may be transported further to the east of the ridge axis owing to the westward ridge migration.

Upwelling at depths shallower than 100 km is intensive and symmetric beneath the ridge axis, in agreement with the prominent zone of silicate melt indicated by the high conductivity in our model (Fig. 3b). Despite the smoothing penalty applied during the data inversion, our model shows a rapid resistivity increase that coincides with the sloping boundaries of the upwelling mantle. Along this boundary, the flow pattern turns from upwelling to horizontal flow as the mantle begins to be dragged off-axis. Because the high resistivity agrees with that of melt-free dry olivine²⁷, this transition away from the inferred melting region is consistent with the closure or compaction of the porous melt network generated by upwelling. This implies that buoyant melt is focused to the ridge axis by the shape of the porous melting region, rather than by upward migration to the base of the much shallower thermal lithosphere defined by the half-space cooling model (Fig. 3b).

The shallow uppermost mantle above 20 km depth is predominantly resistive, suggesting it is largely unmolten with a negligible porosity. However, moderately low vertical resistivity spanning from the ridge axis to 10 km to the east is consistent with a vertical melt channel connecting from the deeper melting region to the base of the crust (Fig. 2). This feature occurs where seismic tomography found a low-velocity anomaly directly beneath the crust¹². Our model suggests that the off-axis velocity anomaly is due to a shallow melt transport channel or network of channels¹³ rather than reflecting deeper upwelling asymmetry. Highly porous transport channels could allow for rapid magma ascent to the crust¹³. This may explain why the high conductivity imaged at depths >20 km does not extend all the way to the base of the crust beneath the ridge axis. Magma generated in the upwelling mantle may reach a critical depth near the top of the melting region where it is efficiently transported to the crust through porous channels that arise episodically in space and time.

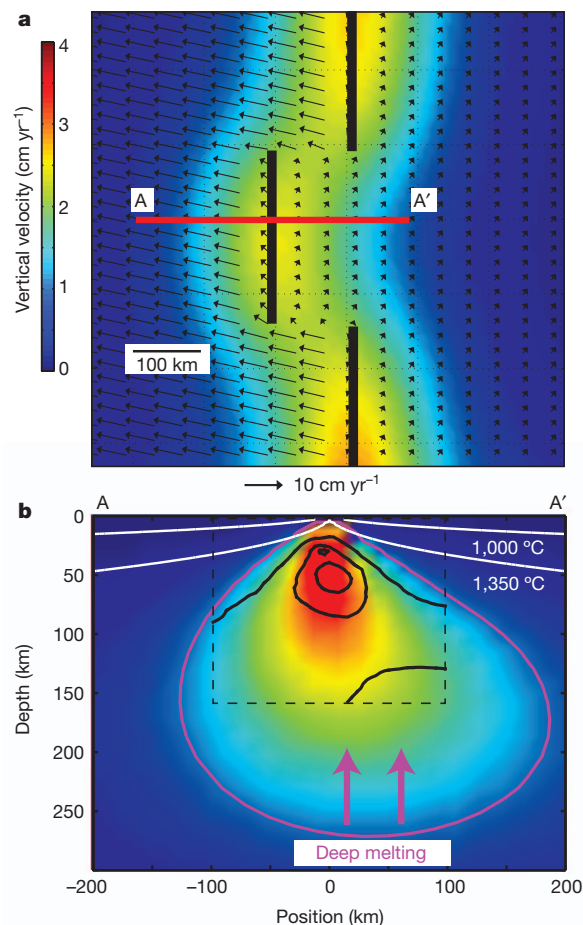


Figure 3 | Three-dimensional passive-flow simulation of the EPR kinematics, including ridge migration, sea-floor spreading and the Clipperton and Siqueiros transform offsets. **a**, Map view showing the imposed absolute plate motions of the Pacific and Cocos plates (black arrows) and the simulated upwelling vertical velocity at 150 km depth (shaded colours). **b**, Cross-section of vertical velocity along profile A–A' (in **a**), which crosses the ridge axis at the same position as the magnetotelluric profile. The magenta line contours 1 cm yr^{-1} vertical velocity, highlighting the deep asymmetry to the east (right). Black lines delineate 0.1, 1 and 10 $\Omega \text{ m}$ resistivity contours from the magnetotelluric inversion model (Fig. 2). White lines show isotherms for a half-space cooling model.

METHODS SUMMARY

We used a robust array processing method²⁸ to estimate magnetotelluric responses from the raw time series and inverted the responses for triaxially anisotropic conductivity using nonlinear two-dimensional inversion implemented with an adaptive finite element algorithm^{29,30}. Flow modelling of mantle upwelling due to absolute plate motions was accomplished using a finite element approach.

Full Methods and any associated references are available in the online version of the paper.

Received 16 October 2012; accepted 23 January 2013.

1. Forsyth, D. W. *et al.* Imaging the deep seismic structure beneath a mid-ocean ridge; the MELT experiment. *Science* **280**, 1221–1224 (1998).
2. Katz, R. F. Porosity-driven convection and asymmetry beneath mid-ocean ridges. *Geochim. Geophys. Geosyst.* **11**, Q0AC07 (2010).
3. Hammond, W. C. & Toomey, D. R. Seismic velocity anisotropy and heterogeneity beneath the Mantle Electromagnetic and Tomography Experiment (MELT) region of the East Pacific Rise from analysis of P and S body waves. *J. Geophys. Res.* **108**, 2176, <http://dx.doi.org/10.1029/2002JB001789> (2003).
4. Dunn, R. A. & Forsyth, D. W. Imaging the transition between the region of mantle melt generation and the crustal magma chamber beneath the southern East Pacific Rise with short-period Love waves. *J. Geophys. Res.* **108**, 2352, <http://dx.doi.org/10.1029/2002JB002217> (2003).

5. Baba, K., Chave, A. D., Evans, R. L., Hirth, G. & Mackie, R. L. Mantle dynamics beneath the East Pacific Rise at 17°S: insights from the Mantle Electromagnetic and Tomography (MELT) experiment. *J. Geophys. Res.* **111**, B02101, <http://dx.doi.org/10.1029/2004JB003598> (2006).
6. Conder, J. A., Forsyth, D. W. & Parmentier, E. M. Asthenospheric flow and asymmetry of the East Pacific Rise, MELT area. *J. Geophys. Res.* **107**, 2344, <http://dx.doi.org/10.1029/2001JB000807> (2002).
7. Conder, J. A., Forsyth, D. W. & Parmentier, E. M. Asymmetric mantle dynamics in the MELT region of the East Pacific Rise. *Earth Planet. Sci. Lett.* **200**, 287–295 (2002).
8. Wang, Y., Forsyth, D. W. & Savage, B. Convective upwelling in the mantle beneath the Gulf of California. *Nature* **462**, 499–501 (2009).
9. Hirschmann, M. M., Tenner, T., Aubaud, C. & Withers, A. C. Dehydration melting of nominally anhydrous mantle: the primacy of partitioning. *Phys. Earth Planet. Inter.* **176**, 54–68 (2009).
10. Ni, H., Keppeler, H. & Behrens, H. Electrical conductivity of hydrous basaltic melts: implications for partial melting in the upper mantle. *Contrib. Mineral. Petrol.* **162**, 637–650 (2011).
11. Yoshino, T., McIsaac, E., Laumonier, M. & Katsura, T. Electrical conductivity of partial molten carbonate peridotite. *Phys. Earth Planet. Inter.* **194–195**, 1–9 (2012).
12. Toomey, D. R., Jousset, D., Dunn, R. A., Wilcock, W. S. D. & Detrick, R. S. Skew of mantle upwelling beneath the East Pacific Rise governs segmentation. *Nature* **446**, 409–414 (2007).
13. Katz, R. F. & Weatherley, S. M. Consequences of mantle heterogeneity for melt extraction at mid-ocean ridges. *Earth Planet. Sci. Lett.* **335–336**, 226–237 (2012).
14. Aharonov, E., Whitehead, J. A., Kelemen, P. B. & Spiegelman, M. Channeling instability of upwelling melt in the mantle. *J. Geophys. Res.* **100** (B10), 20433–20450 (1995).
15. Key, K. & Constable, S. Broadband marine MT exploration of the East Pacific Rise at 9°50'N. *Geophys. Res. Lett.* **29**, 2054, <http://dx.doi.org/10.1029/2002GL016035> (2002).
16. Gripp, A. E. & Gordon, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150**, 321–361 (2002).
17. Haymon, R. M. *et al.* Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at 9°45'–52'N: direct submersible observations of seafloor phenomena associated with an eruption event in April 1991. *Earth Planet. Sci. Lett.* **119**, 85–101 (1993).
18. Baker, E. T. *et al.* Hydrothermal plumes along the East Pacific Rise, 8°40' to 11°50'N: plume distribution and relationship to the apparent magmatic budget. *Earth Planet. Sci. Lett.* **128**, 1–17 (1994).
19. Kent, G. M., Harding, A. J. & Orcutt, J. A. Distribution of magma beneath the East Pacific Rise between the Clipperton Transform and the 9°17'N Deval from forward modeling of common depth point data. *J. Geophys. Res.* **B 98**, 13945–13969 (1993).
20. Tolstoy, M. *et al.* A seafloor spreading event captured by seismometers. *Science* **314**, 1920–1922 (2006).
21. Danyushevsky, L. V., Eggins, S. M., Falloon, T. J. & Christie, D. M. H₂O abundance in depleted to moderately enriched mid-ocean ridge magmas; part I: incompatible behaviour, implications for mantle storage, and origin of regional variations. *J. Petrol.* **41**, 1329–1364 (2000).
22. Saal, A. E., Hauri, E. H., Langmuir, C. H. & Perfit, M. R. Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle. *Nature* **419**, 451–455 (2002).
23. Hirose, K. Partial melt compositions of carbonated peridotite at 3 GPa and role of CO₂ in alkali-basalt magma generation. *Geophys. Res. Lett.* **24**, 2837–2840 (1997).
24. Wendlandt, R. F. & Mysen, B. O. Melting phase relations of natural peridotite + CO₂ as a function of degree of partial melting at 15 and 30 kbar. *Am. Mineral.* **65**, 37–44 (1980).
25. Dasgupta, R., Hirschmann, M. M. & Smith, N. D. Water follows carbon: CO₂ incites deep silicate melting and dehydration beneath mid-ocean ridges. *Geology* **35**, 135–138 (2007).
26. Minarik, W. G. Complications to carbonate melt mobility due to the presence of an immiscible silicate melt. *J. Petrol.* **39**, 1965–1973 (1998).
27. Constable, S. SEO3: A new model of olivine electrical conductivity. *Geophys. J. Int.* **166**, 435–437 (2006).
28. Egbert, G. D. Robust multiple-station magnetotelluric data processing. *Geophys. J. Int.* **130**, 475–496 (1997).
29. Key, K. & Oval, J. A parallel goal-oriented adaptive finite element method for 2.5-D electromagnetic modelling. *Geophys. J. Int.* **186**, 137–154 (2011).
30. Constable, S. C., Parker, R. L. & Constable, C. G. Occam's inversion — a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics* **52**, 289–300 (1987).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the captain and crew of R/V *Roger Revelle*, members of the scientific party (J. Behrens, G. Boran, G. Heinson, J. King, A. Massarweh, L. Terzi and C. Weiss), and the technical team (C. Armerding, P. Cheng, C. Berger, G. Englehorn, G. Howe, K. Callaway and J. Lemire). We thank D. Myer for assistance with the data processing and R. Katz for reviewing the manuscript. This work was funded by US NSF award OCE-0241597 and the Seafloor Electromagnetic Methods Consortium at Scripps Institution of Oceanography.

Author Contributions K.K. and S.C. collected and analysed the field data. A.P. and K.K. performed the melt calculations. L.L. carried out the mantle flow simulations. K.K. wrote most of the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.K. (kkey@ucsd.edu).

METHODS

Magnetotelluric data analysis. The 29 broadband magnetotelluric stations were deployed during a single month-long cruise, yielding frequency domain magnetotelluric responses in the 20–5,000 s period band. The station names are given in Supplementary Fig. 1 and the magnetotelluric data are shown in Supplementary Figs 2 and 3. Magnetotelluric impedance skews³¹ for each station are displayed in Supplementary Fig. 4. Low skews from the ridge axis to the west are consistent with a predominantly two-dimensional conductivity beneath the ridge. Many stations on the Cocos plate have large impedance skews at periods of 100–1,000 s, indicating three-dimensional conductivity effects. These large skews are probably due to the 85 km lateral offset of ridge axis, and by inference the conductive upwelling region, across the Clipperton transform, which is located only 75 km north of the magnetotelluric profile. For our two-dimensional analysis, we omitted all transverse-electric (TE) data (that is, ridge-parallel electric fields) from sites s48 to s60 and only included transverse magnetic (TM) data (that is, ridge-perpendicular vertical current loops), because TM data are considered more reliable for two-dimensional interpretation in the presence of three-dimensional effects^{31,32}.

Nonlinear anisotropic inversion. We modelled the data using a regularized nonlinear inversion that solved for anisotropic electrical resistivity with a triaxial form:

$$\rho = \begin{bmatrix} \rho_x & 0 & 0 \\ 0 & \rho_y & 0 \\ 0 & 0 & \rho_z \end{bmatrix} \quad (1)$$

where the x direction is aligned with the ridge axis, y is ridge perpendicular and z points down. The inversion solved for the three unknown resistivity tensor elements on a grid of 7,728 triangular cells spanning from the sea floor into the mantle, with small cells conforming to the rugged sea-floor bathymetry and larger cells at depth. The total number of unknown parameters was 23,184.

The inversion follows the usual Occam's approach by minimizing the following functional³⁰:

$$U = \|\hat{m}\|^2 + \mu^{-1} \|W(d - F(m))\|^2 \quad (2)$$

The first term measures the model roughness. The second term measures the fit of the model m (\log_{10} resistivity) to the observed data d , weighted by the inverse data standard errors in the diagonal matrix W , where F is the forward modelling operator and $\|\dots\|$ is the two-norm. Occam uses an automatic search method to find the largest trade-off parameter μ that fits the data, thereby seeking the smoothest model fitting the data. For the triaxially anisotropic resistivity considered here, the model roughness term is expanded into the form:

$$\|\hat{m}\|^2 = \|\hat{m}_x\|^2 + \|\hat{m}_y\|^2 + \|\hat{m}_z\|^2 + \|m_x - m_y\|^2 + \|m_y - m_z\|^2 + \|m_z - m_x\|^2 \quad (3)$$

The first three terms measure the spatial roughness of each anisotropic resistivity component by differencing the values for adjacent resistivity parameters. The last three terms measure the degree of anisotropy within each cell so that the inversion only generates anisotropy where it is required to fit the observed data. Equation (3) has an implicit weighting of unity between the spatial roughness and anisotropic roughness measures so that neither is given preferential treatment by the inversion.

We applied a 10% error floor to the data and then ran the inversion until it found a minimum root-mean-squared (r.m.s.) misfit of 1.73. We then restarted the inversion and ran it to a target r.m.s. misfit that was 10% larger (1.90), to ensure the data were not being over-fitted. The inversion was started from a uniform 10 Ω m model. Supplementary Fig. 5 shows the resulting anisotropic inversion model and Supplementary Fig. 3 shows the corresponding model fits to the data.

We also carried out an inversion of only the TM mode data to examine differences in the inverted model structure as a function of data constraints (Supplementary Fig. 6). This inversion finds a deep asymmetric conductor below 100 km depth that is similar to the previous inversion of both the low-skew TE mode data and all TM data (Supplementary Fig. 5). The TM-only inversion similarly finds a shallower ridge-centred melting region, but its conductivity is much lower than the previous inversion that included TE data. This is not surprising because the TM mode is known to be less sensitive to the conductivity of two-dimensional conductors than the TE mode, but is more sensitive to the lateral edges of conductors than the TE mode³¹.

Model sensitivity studies. We tested the sensitivity of the data to both the shallow (A) and deep (B) conductors and the sub-crustal vertical conductor (C) by erasing each feature and computing the corresponding change in data fit. Supplementary Fig. 7 shows the outline of these conductive regions, where the resistivity was increased to 100 Ω m (A and B) and 200 Ω m (C). Supplementary Fig. 8 shows the increase in the r.m.s. data misfit for each test. Removing the shallow conductor A resulted in an r.m.s. increase of 400% for the TE data and 20% for the TM data, showing that it is strongly required by the data. Removing the deeper conductor B resulted in a change of only 20% in the TM mode data fit, showing that it is less well constrained by the data, yet still required to obtain a good overall fit. Removing the vertical conductor C resulted in a 3% change in the TM mode fit; while this is a small change, individual sites have up to 9% worse data misfits when the conductor is removed, suggesting it is a necessary feature of the model.

Three-dimensional passive-flow modelling. We constructed a three-dimensional flow model simulating the kinematics of the EPR transform faults near 8° and 10° N using a finite element method³³. The model domain spans 12° (latitude), 24° (longitude) and 1,000 km (depth), where the long enough east–west dimension of the box avoids the edge effects. A high mesh resolution (~7 km) was selected so that the flow pattern does not depend on the grid size. Plate motions were imposed as surface velocity boundary conditions, with a full spreading rate of 11 cm yr⁻¹ and a northwestward ridge migration rate of 5 cm yr⁻¹, according to NUVELIA¹⁶. Sea-floor ages were used to prescribe the upper thermal boundary layer, with zero age at the spreading centre, which increases to 30 Myr towards the edges of the box. A temperature- and depth-dependent viscosity structure was used, with the mean viscosities at depths of lithosphere, upper mantle, transition zone and lower mantle corresponding to 100, 0.01, 1, 30 (relative to 10²¹ Pa s), respectively.

31. Berdichevsky, M. & Dmitriev, V. I. *Models and Methods of Magnetotellurics* (Springer, 2008).
32. Wannamaker, P. E., Hohmann, G. W. & Ward, S. H. Magnetotelluric responses of three-dimensional bodies in layered earths. *Geophysics* **49**, 1517–1533 (1984).
33. Zhong, S., Zuber, M. T., Moresi, L. & Gurnis, M. Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *J. Geophys. Res.* **105**, 11063–11082 (2000).

Tubicolous enteropneusts from the Cambrian period

Jean-Bernard Caron¹, Simon Conway Morris² & Christopher B. Cameron³

Hemichordates are a marine group that, apart from one monospecific pelagic larval form, are represented by the vermiform enteropneusts and minute colonial tube-dwelling pterobranchs. Together with echinoderms, they comprise the clade Ambulacraria¹. Despite their restricted diversity, hemichordates provide important insights into early deuterostome evolution, notably because of their pharyngeal gill slits². Hemichordate phylogeny has long remained problematic^{3,4}, not least because the nature of any transitional form that might serve to link the anatomically disparate enteropneusts and pterobranchs is conjectural. Hence, inter-relationships have also remained controversial. For example, pterobranchs have sometimes been compared to ancestral echinoderms¹. Molecular data identify enteropneusts as paraphyletic, and harrimaniids^{5–8} as the sister group of pterobranchs. Recent molecular phylogenies suggest that enteropneusts are probably basal within hemichordates, contrary to previous views⁹, but otherwise provide little guidance as to the nature of the primitive hemichordate^{7,8}. In addition, the hemichordate fossil record is almost entirely restricted to peridermal skeletons of pterobranchs, notably graptolites^{10,11}. Owing to their low preservational potentials, fossil enteropneusts are exceedingly rare^{12–15}, and throw no light on either hemichordate phylogeny or the proposed harrimaniid–pterobranch transition. Here we describe an enteropneust, *Spartobranchus tenuis* (Walcott, 1911), from the Middle Cambrian-period (Series 3, Stage 5) Burgess Shale. It is remarkably similar to the extant harrimaniids, but differs from all known enteropneusts in that it is associated with a fibrous tube that is sometimes branched. We suggest that this is the precursor of the pterobranch periderm, and supports the hypothesis that pterobranchs are miniaturized and derived from an enteropneust-like worm^{5,6}. It also shows that the periderm was acquired before size reduction and acquisition of feeding tentacles, and that colonality emerged through aggregation of individuals, perhaps similar to the Cambrian rhabdopleurid *Fasciculitubus*¹¹. The presence of both enteropneusts and pterobranchs in Middle Cambrian strata, suggests that hemichordates originated at the onset of the Cambrian explosion.

Phylum Hemichordata Bateson, 1885

Class Enteropneusta Gegenbaur, 1870

Spartobranchus tenuis gen. nov.

Etymology. *Sparte*, Greek for cord or rope, and *branchos* for gills.

Lectotype. USNM 108494

Referred material. Paralectotypes USNM 553526 (see Supplementary Information, section 1), in addition, 5,277 specimens of the Smithsonian Institution National Museum of Natural History (SI-NMNH), 3,608 of the Royal Ontario Museum (ROM) and 128 of the Geological Survey of Canada (GSC).

Locality. Walcott Quarry in Yoho National Park, British Columbia, Canada.

Horizon. Walcott Quarry Member, Burgess Shale Formation, Lower Chancellor Group.

Diagnosis. Vermiform, maximum length 10 cm. Body is comprised of a short proboscis, collar, pharyngeal area with up to 40 pairs of gill and

tongue bars, and elongate posterior trunk ending in a bulbous unit. Gut straight, anus terminal. Approximately one-quarter of the specimens are associated with a tube that has organic walls. This tube varies from straight, completely circular to helicoidal, and sometimes branching. No more than one individual per tube has been observed.

S. tenuis was originally classified as *Ottoia tenuis*, Walcott, 1911 (ref. 16). Conway Morris¹⁷ removed this taxon from *Ottoia* (type species *Ottoia prolifica* (Walcott¹⁶)), which is a priapulid¹⁸, and in passing drew attention to its similarity to enteropneusts. Specimens of *S. tenuis* are preserved as flat carbonaceous compressions. They often occur in large clusters (Supplementary Information, sections 2 and 4), and this taxon is one of the most abundant species in the Walcott Quarry community^{19,20}. A possible related form from the Middle Cambrian Wheeler Formation, Utah, United States (*Spartobranchus* sp.; Supplementary Information, section 3).

The body is flexible, consisting of a short proboscis, collar and narrow elongate trunk terminating in a bulbous structure. The largest complete individuals are approximately 10 cm long with the proboscis accounting for less than 5% of the total body length. The trunk width represents less than 2% of the total length. The proboscis is usually rounded (Figs 1a, b and 2a–c), triangular (Figs 1c, d, f, g, i–l and 2d–e, h–j) or, more rarely, elongate (Fig. 1e). Internal features of the proboscis include a darker axial zone (Fig. 1a, e, f) and a posterior crescent structure (Fig. 1b, d). These are consistent with the position of the kidney–stomochord complex and a pre-oral ciliary organ, respectively. From the proboscis a thin triangular element also extends posteriorly and dorsally towards the collar. This structure is most evident when the proboscis is disarticulated and rotated (Fig. 1g). It connects the proboscis to the rest of the trunk and is presumably related to the stomochord. An additional Y-shaped structure connects the proboscis to the collar region in the same area and may represent the nuchal skeleton (Fig. 1b, c, f, g–i). The mouth is located on the anterior ventral face of the collar (Fig. 1c, d, f). The short collar is usually poorly preserved, and appears as an inflated area around the anterior part of the trunk (Figs 1c, e and 2h, i). The pharyngeal region accounts for approximately 10–20% of the total length, and it is wider towards the mouth and tapers posteriorly (Figs 1a, f, k and 2a–e, h–j). The dorsal branchial pharynx comprises thick and thin bars, interpreted as gill and tongue bars, respectively. In total there are up to 40 pairs, and these are connected symmetrically along the dorsal surface but with no evidence of synapicalae joining the bars (Fig. 1a, f, g, k, l). Darker structures on the dorsal and ventrolateral sides of the branchial pharynx may represent an epibranchial ridge (Fig. 1f, g, k, l) and parabanchial ridges, respectively; the latter serve to demarcate the dorsal branchial from the ventral digestive pharynx (Fig. 1f, i–k). These ridges merge posteriorly into an ovoid structure, interpreted as the oesophageal organ (Figs 1a, f, k and 2a, d, e, h, j). The trunk is smooth and featureless, and its width mostly defined by the width of the gut. The terminus is bulbous (Figs 1a and 2a–f, h, i, and Supplementary Information 4d, f) and can be as wide as the proboscis. The gut is straight and the anus is terminal (Fig. 2d).

Approximately one-quarter of specimens are associated with tubes that have a fibrous composition (Fig. 2h, j). They vary from straight,

¹Department of Natural History (Palaeobiology), Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada. ²Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK. ³Département de sciences biologiques, Université de Montréal, Pavillon Marie-Victorin, Montréal QC H2V 2S9, Canada.

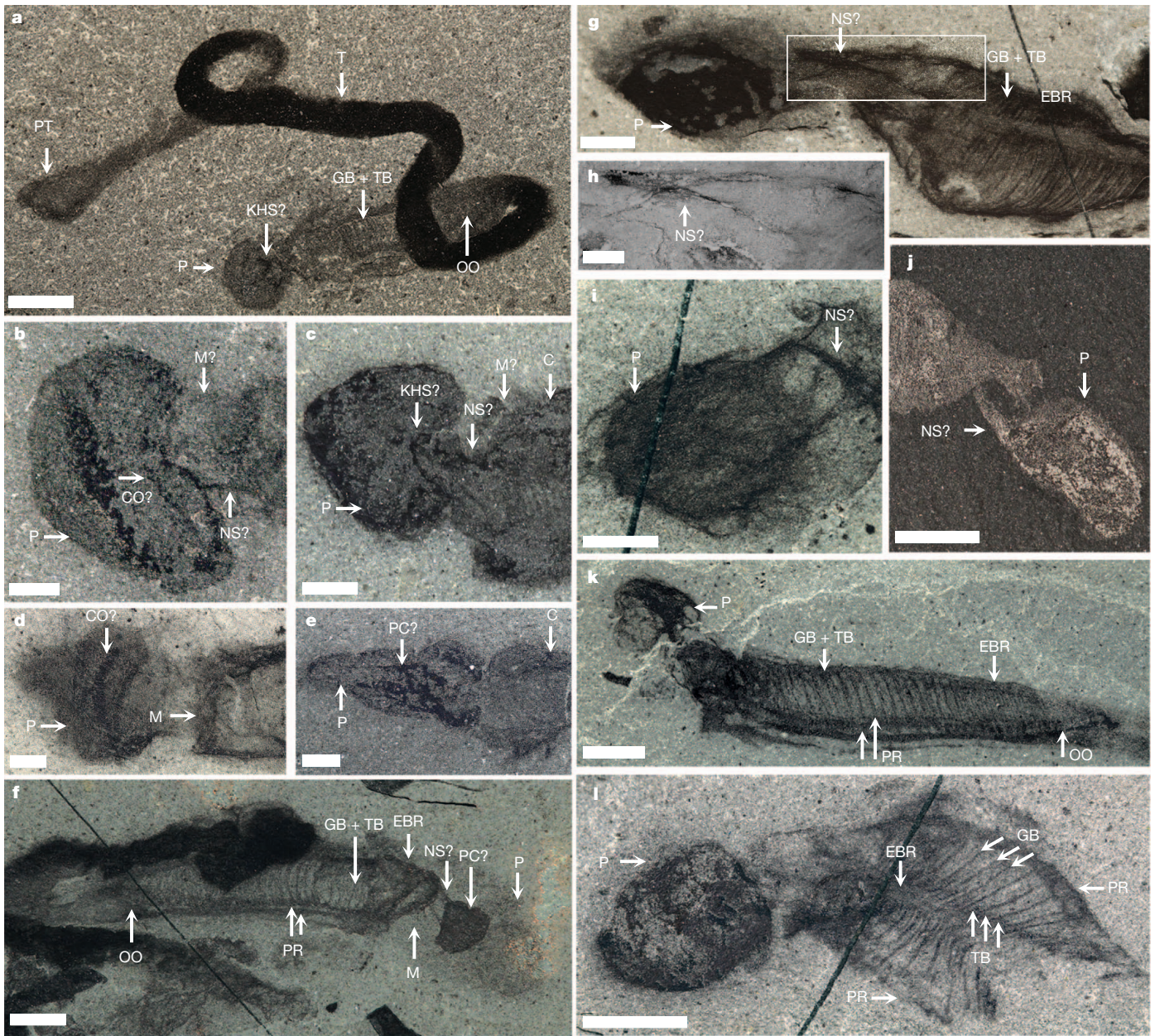


Figure 1 | *Spartobranchus tenuis* (Walcott, 1911) individuals from the Burgess Shale. **a**, Complete specimen, ROM 62123. **b–e**, Close-up views of the anterior section. **b**, Lateral view, USNM 202861. **c**, Oblique view, GSC 136961. **d**, Ventral view, ROM 62124. **e**, Oblique view, USNM 202781. **f**, Lateral view, USNM 202183/202493 (see also Supplementary Information, section 2b, c). **g**, Oblique view, ROM 62125. **h**, Scanning electron microscopy image of the framed area in **g**, ROM 62125. **i**, Oblique view, ROM 62126. **j**, Lateral view,

USNM 202472. **k**, Lateral view, ROM 62127. **l**, Ventral view, USNM 202780. C, collar; CO?, pre-oral ciliary organ; EBR, epibranchial ridge; GB, gill bars; KHS?, kidney–heart–stomochord complex; M (or M?), mouth; NS?, nuchal skeleton; OO, oesophageal organ; P, proboscis; PC?, proboscis coelom; PR, parabranchial ridges; PT, posterior trunk; T, trunk; TB, tongue bars. Question marks indicate uncertainty. Scale bars represent 2 mm (**a**, **f**, **j**, **k**, **l**), 1 mm (**b**, **c**, **e**, **g**), and 0.5 mm (**d**, **h**, **i**). See also Supplementary Information, section 4.

completely circular, branching to helicoidal (Fig. 2). No more than one individual per tube has been observed, and it is evident from the varied (Fig. 2) positions in which the worm was buried that it was not permanently attached. These range from being centred (Fig. 2b, c), located close to one end (Fig. 2a, d, e), to being fully extended (Fig. 2b, c), or contorted or twisted (Fig. 2d–f). Variation in both tube shape and disposition of the occupant suggests that the association was comparatively free, although the bulbous posterior may have been used as an anchor. In times of stress the worms may have vacated their tubes (or may conceivably have produced tubes). Specimens in which the proboscis is in close proximity to the terminal trunk (Fig. 2h, i) may represent a quiescent state. The close match between proboscis and tube when the worm is at one end of the burrow (Fig. 2d, e, j) suggests

that the proboscis was responsible for tube construction. Tubes are often corrugated (Fig. 2b, c, f), presumably reflecting local thickening to impart strength. It is likely that the tube of *S. tenuis* provided a stable shelter, and that the side-branches were used for escape (Supplementary Information, section 5). The absence of obvious sediment in the gut suggests that *S. tenuis* may have pumped water using its ciliated gill slits.

The relationship between hemichordates and echinoderms as ambulacrarians is well supported^{1,3–8}, and hemichordates continue to have a critical role in early deuterostome and chordate evolution studies. Nevertheless, although gill slits are homologous², other features (for example, the stomochord and post-anal tail) that were once thought to be shared with chordates are probably convergent^{21,22}. These uncertainties

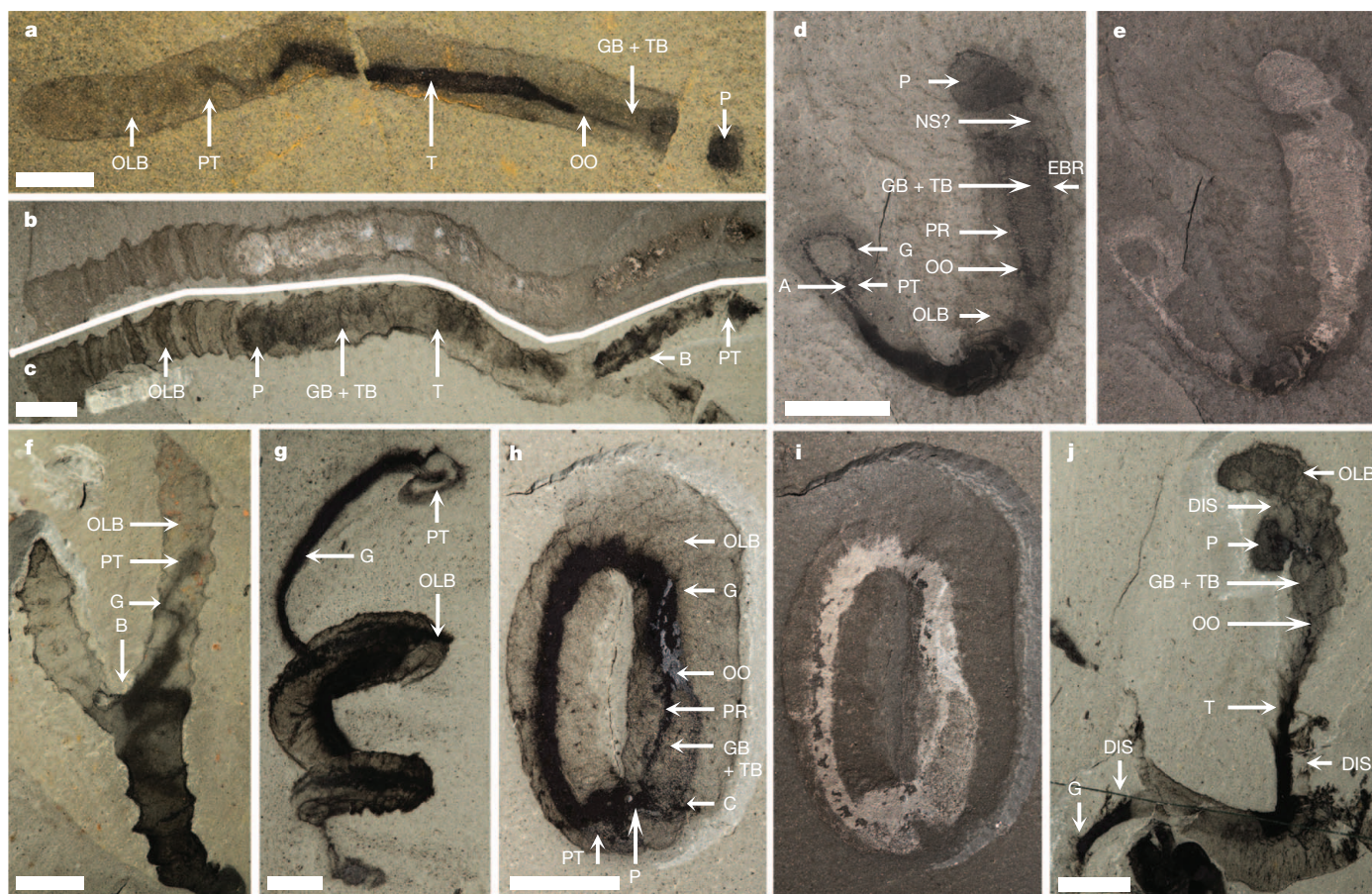


Figure 2 | *Spartobranchus tenuis* (Walcott, 1911) individuals associated with tubular structures, from the Burgess Shale. **a**, Linear tube, USNM 203032. **b**, **c**, Linear and branching tube, USNM 509820. **d**, **e**, U-shaped tube, USNM 509806/509814. **f**, Branching tube, ROM 57170. **g**, Helicoidal tube,

ROM 62128. **h**, **i**, Circular tube, USNM 275692. **j**, Partially disassociated tube, ROM 62129; see also Supplementary Information, section 5a. A, anus; B, branching structure; DIS, disassociated; G, gut; OLB, organic-lined burrow. Scale bars represent 5 mm (**a–f**, **h–j**) and 2.5 mm (**g**).

serve to amplify yet further the difficulties in delineating the nature of the urdeuterostome. The notably disparate anatomy of enteropneusts and pterobranchs has also led to disagreement as to whether very early deuterostomes were solitary and motile^{5,6}, or colonial and semi-sessile⁹. The idea that they were solitary and motile is more consistent with the development of the harrimaniid *Saccoglossus*^{2,23,24}, but more recent molecular analyses tend to identify both pterobranchs and harrimaniids either in a basal clade⁷ or in a basal polytomy⁸ within hemichordates. In the absence of a clear phylogenetic signal, the morphology of the primitive hemichordate must remain open to question. In principle, fossils might provide direct evidence, but the available record is strongly biased in favour of pterobranchs, including graptolites^{10,11}, whose collagenous periderm has a high preservation potential. They are an important component of early to middle Palaeozoic-era marine biotas. In contrast, the soft zooids are very seldom preserved and shed no light on hemichordate origins²⁵, and given their soft-bodied anatomy, the fossil record of enteropneusts is exceptionally meager^{12–15} and restricted to occurrences in Konservat-Lagerstätten. Indirect evidence from trace fossils is mostly equivocal^{26,27}, although a Triassic-period record²⁸ is consistent with an early invasion of the deep sea by torquaratorid enteropneusts⁸.

Here we provide direct evidence for a Cambrian enteropneust, and thus extend their known range by approximately 200 million years. This discovery has two important implications. In *S. tenuis*, the absence of synapticulae between the gill and tongue bars, and of hepatic sacs on the trunk, point to affinities with present-day harrimaniid worms. A generic relationship is more difficult to resolve because several key soft tissue characteristics of the internal anatomy are absent or not preserved, including details of the collar nerve cord, coelomic

diverticula, gonads or a juvenile post-anal tail. Two lines of evidence suggest that the collagenous coenecium of pterobranchs had origins in the organic tubes of *S. tenuis*. First, the tubes of *S. tenuis* have a fibrous structure that sometimes shows evidence of tearing. We have no evidence of characteristic growth bands (fuselli), a pterobranch innovation that probably appeared with a reduction in size and the evolution of a shield-shaped secretory organ. Second, the secretory proboscis and cephalic shield are homologous in enteropneust and pterobranchs (both protosoma), and thus the transition from inferred involvement of the proboscis in tube construction by *S. tenuis* to the secretory role of the cephalic shield is unproblematic. We suggest that in the lineage leading to the Enteropneusta, the tube lining was lost, whereas in pterobranchs it provided a substrate for an astonishing radiation, not least in Palaeozoic planktonic ecosystems.

In conclusion, our study supports the presence of tubicolous harrimaniid enteropneusts in the Cambrian, and provides direct evidence for a link between harrimaniids and pterobranchs. Whether *S. tenuis* represents a stem-group harrimaniid, a stem-group pterobranch or a stem-group of pterobranch plus harrimaniid, or even a stem group hemichordate, is not certain. Moreover, it is worth noting that any morphological differences between these now distinct clades may have been minimal in the Cambrian. However, as *S. tenuis* seems to be closer to the harrimaniids than to any other hemichordates, including pterobranchs, a stem-group pterobranch seems unlikely. This would suggest that tube construction pre-dated the shift from an harrimaniid enteropneust ancestor to the tentaculate pterobranch and that tubes were secondarily lost in enteropneusts. Details of the shift to a miniaturized form and coloniality await the discovery of further well-preserved material, but Cambrian pterobranchs with the simplified

fuselli may provide possible candidates¹¹. If the primitive position of harrimaniids and pterobranchs is confirmed, then the 'dipleurula' larvae that characterize ptychoderids and echinoderms may be a result of homoplasy or secondary loss in the harrimaniid clade. In addition, some chordate features that characterize ptychoderids, including gill bar synaptaculæ and a dorsal nerve cord with a hollow canal (in a few species), may be homoplasies. In enteropneusts these characters are correlated with large size, and given the present findings, were probably absent in their common ancestor with the chordates.

METHODS SUMMARY

Sediment that coated sections of some specimens was removed mechanically using a micro-engraving tool equipped with a carbide bit. Specimens were photographed under different illuminations, and cross-polarized light techniques usually provided the best contrasts. Backscatter scanning electron microscopy images of uncoated specimens were obtained of fine anatomical details. Pictures of both parts and counterparts were merged in Adobe PhotoshopCS6 using the 'Apply Image' function and 'Darken' blending mode (Fig. 1b, d, f, g, i, k, l).

Received 2 December 2012; accepted 18 February 2013.

Published online 13 March; corrected online 27 March 2013 (see full-text HTML version for details).

- Swalla, B. J. & Smith, A. B. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. *Phil. Trans. R. Soc. B* **363**, 1557–1568 (2008).
- Gonzalez, P. & Cameron, C. B. The gill slits and pre-oral ciliary organ of *Protoglossus* (Hemichordata: Enteropneusta) are filter-feeding structures. *Biol. J. Linn. Soc.* **98**, 898–906 (2009).
- Winchell, C. J., Sullivan, J., Cameron, C. B., Swalla, B. J. & Mallatt, J. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol. Biol. Evol.* **19**, 762–776 (2002).
- Cameron, C. B. A phylogeny of the hemichordates based on morphological characters. *Can. J. Zool.* **83**, 196–215 (2005).
- Cannon, J. T., Rychel, A. L., Eccleston, H., Halanych, K. M. & Swalla, B. J. Molecular phylogeny of Hemichordata, with updated status of deep-sea enteropneusts. *Mol. Phylogenet. Evol.* **52**, 17–24 (2009).
- Cameron, C. B., Garey, J. R. & Swalla, B. J. Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc. Natl Acad. Sci. USA* **97**, 4469–4474 (2000).
- Worsaae, K., Sterrer, W., Kaul-Strehlow, S., Hay-Schmidt, A. & Giribet, G. An anatomical description of a miniaturized acorn worm (Hemichordata, Enteropneusta) with asexual reproduction by paratomy. *PLoS ONE* **7**, e48529 (2012).
- Osborn, K. J. *et al.* Diversification of acorn worms (Hemichordata, Enteropneusta) revealed in the deep sea. *Proc. R. Soc. B* **279**, 1646–1654 (2012).
- Hyman, L. H. in *The Invertebrates*. Vol. 5, 72–207 (McGraw-Hill, 1959).
- Mitchell, C. E., Melchin, M. J., Cameron, C. B. & Maletz, J. Phylogenetic analysis reveals that *Rhabdopleura* is an extant graptolite. *Lethaia* **46**, 34–56 (2012).
- Rickards, R. B. & Durman, P. N. in *Studies in Palaeozoic Palaeontology* National Museum of Wales Geological Series Vol. 25 (eds Bassett, M. G. & Deisler, V. K.) 5–92 (National Museum of Wales, 2006).
- Bardack, D. in *Richardson's guide to the fossil fauna of Mazon Creek* (eds Shabica, C. W. & Hay, A. A.) 89–92 (Northeastern Illinois University, 1997).
- Arduini, P., Pinna, G. & Terruzzi, G. *Megaderia sinemuriense* n.g. n.sp., a new fossil enteropneust of the Sinemurian of Osteno in Lombardy. *Atti della Società Italiana di Scienze Naturali e del Museo Civico di Storia Naturale di Milano* **122**, 104–108 (1981).
- Alessandro, A., Bracchi, G. & Riou, B. Polychaete, sipunculan and enteropneust worms from the Lower Callovian (Middle Jurassic) of La Voulte-sur-Rhône (Ardèche, France). *Memoire della Società Italiana di Scienze Naturali e del Museo Civico di Storia Naturale di Milano (Fascicolo I)* **32**, 1–16 (2004).
- Bechly, G. & Frickhinger, K. A. in *The fossils of Solnhofen 2*. 76–79 (Goldschnecks-Verlag, 1999).
- Walcott, C. Cambrian Geology and Paleontology II. Middle Cambrian annelids. *Smiths. Misc. Collect.* **57**, 109–145 (1911).
- Conway Morris, S. The Burgess Shale (Middle Cambrian) fauna. *Annu. Rev. Ecol. Syst.* **10**, 327–349 (1979).
- Conway Morris, S. Fossil priapulid worms. *Special Papers in Palaeontology* **20**, 1–95 (1977).
- Conway Morris, S. The community structure of the Middle Cambrian phyllopod bed (Burgess Shale). *Palaeontology* **29**, 423–467 (1986).
- Caron, J.-B. & Jackson, D. A. Paleoeology of the Greater Phyllopod Bed community, Burgess Shale. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **258**, 222–256 (2008).
- Stach, T. & Kaul, S. The postanal tail of the enteropneust *Saccoglossus kowalevskii* is a ciliary creeping organ without distinct similarities to the chordate tail. *Acta Zool.* **92**, 150–160 (2011).
- Ruppert, E. E. Key characters uniting hemichordates and chordates: homologies or homoplasies? *Can. J. Zool.* **83**, 8 (2005).
- Pani, A. M. *et al.* Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289–294 (2012).
- Lowe, C. J. *et al.* Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
- Rickards, R. B. & Stait, B. A. *Psigraptus*, its classification, evolution and zooid. *Alcheringa* **8**, 101–111 (1984).
- Kaźmierczak, J. & Pszczółkowski, A. Burrows of Enteropneusta in Muschelkalk (Middle Triassic) of the Holy Cross Mountains, Poland. *Acta Palaeontol. Pol.* **14**, 299–318 (1969).
- Wetzel, W. Flintgefüllte enteropneusten — wohnröhren in einen oberkreide-Geschiebe. *Schriften des Naturwissenschaftlichen Vereins für Schleswig-Holstein* **42**, 104–107 (1972).
- Twitchett, R. J. The resting trace of an acorn-worm (Class: Enteropneusta) from the Lower Triassic. *J. Paleontol.* **70**, 128–131 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Lowe for constructive comments, D. Erwin and M. Florence for access to the Smithsonian Institution collections, Parks Canada for collection and research permits, P. Fenton and V. Brown for assistance, R. Gaines for donation of specimens from Utah, and M. Collins (ARTofFACT) for reconstructions. J.-B.C. and C.B.C. are supported by Natural Sciences and Engineering Research Council Discovery Grants, and S.C.M. is supported by St John's College, University of Cambridge, UK. This is Royal Ontario Museum Burgess Shale project number 41.

Author Contributions All authors contributed to write the main paper and commented on the manuscript at all stages. J.-B.C. conducted the observations and photography of specimens, prepared figures and wrote the Supplementary Information.

Author Information The LSIDs urn:lsid:zoobank.org:pub:BCEFDE43-8319-4C02-A184-EF5995814B86 (article) and urn:lsid:zoobank.org:act:6EC4624D-BB51-45D5-A4AA-FC48BA92FE42 (taxon) have been deposited in ZooBank. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-B.C. (jcaron@rom.on.ca).

Preservation of ovarian follicles reveals early evolution of avian reproductive behaviour

Xiaoting Zheng^{1,2}, Jingmai O'Connor³, Fritz Huchzermeyer⁴, Xiaoli Wang¹, Yan Wang¹, Min Wang³ & Zhonghe Zhou³

The two groups of archosaurs, crocodilians and birds, form an extant phylogenetic bracket for understanding the reproductive behaviour of dinosaurs. This behaviour is inferred from preserved nests and eggs, and even gravid individuals¹. Data indicate that many 'avian' traits were already present in Paraves—the clade that includes birds and their close relatives—and that the early evolution of the modern avian form of reproduction was already well on its way^{2,3}. Like living neornithine birds, non-avian maniraptorans had daily oviposition and asymmetrical eggs with complex shell microstructure, and were known to protect their clutches^{4–6}. However, like crocodilians, non-avian maniraptorans had two active oviducts (one present in living birds), relatively smaller eggs, and may not have turned their eggs in the way that living birds do^{1,6}. Here we report on the first discovery of fossilized mature or nearly mature ovarian follicles, revealing a previously undocumented stage in dinosaur reproduction: reproductively active females near ovulation. Preserved in a specimen of the long bony-tailed *Jeholornis* and two enantiornithine birds from the Early Cretaceous period lacustrine Jehol Biota in northeastern China, these discoveries indicate that basal birds only had one functional ovary, but retained primitive morphologies as a result of their lower metabolic rate relative to living birds. They also indicate that basal birds reached sexual maturity before skeletal maturity, as in crocodiles and paravian dinosaurs. Differences in follicular morphology between *Jeholornis* and the enantiornithines are interpreted as forming an evolutionary gradient from the reproductive condition in paravian dinosaurs towards neornithine birds. Furthermore, differences between the two enantiornithines indicate that this lineage might also have evolved advanced reproductive traits in parallel to the neornithine lineage.

Three specimens of Mesozoic birds from the Tianyu Natural History Museum, Shandong, China (STM), have been found to contain mature ovarian follicles (a follicle being the primary oocyte and the sac in which it is contained⁷) preserved in their body cavity. The largest specimen (STM2-51 slab and counterslab; Fig. 1 and Supplementary Fig. 1) is referable to *Jeholornis* sp., a basal bird with a bony tail longer than that of *Archaeopteryx* (see Supplementary Information). The specimen, in dorsoventral view, preserves approximately 20 follicles ventral to the vertebral column and cranial to the pelvic girdle, overlapped by thoracic ribs. The follicles are all round and subequal in size, ranging from approximately 7.1 to 8.8 mm in diameter (Fig. 1c, d, Supplementary Fig. 2 and Supplementary Table 1); the total number of follicles may have been higher than that observed in all specimens, obscured by overlap. In STM2-51, the follicles are distributed in rows between the caudal end of the sternum and the pelvis; the number of follicles in each row decreases caudally (three to four proximally, one to two distally); however, this pattern may be exaggerated by a crack that laterally borders the preserved follicles. In the main slab, the follicles are preserved permineralized by a dark macrocrystalline mineral (too thick to be eggshell; Fig. 1d). In the counterslab, the follicles are

preserved as pale pink impressions, although some small crystals of the black mineral are also present (Fig. 1c).

The two smaller specimens (STM29-8 and STM10-45; Fig. 2 and Supplementary Figs 3–7) are referable to Enantiornithes (see Supplementary Information); this large clade forms the sister group to Ornithomorphs, the clade that includes living birds (Neornithes), and represents the first major avian radiation^{8,9}. The two specimens vary in proportions of the humerus/ulna and femur/tibiotarsus and the shape of the coracoid, indicating that they represent different taxa, although poor preservation prevents identifying either specimen at the species level (see Supplementary Information). Both specimens preserve mature ovarian follicles in their body cavity visible in dorsoventral (slab and counterslab) view.

Enantiornithes indeterminate (indet.) STM29-8 (Fig. 2c and Supplementary Fig. 5) preserves at least 12 follicles *in situ* along the left side of the vertebral column next to the synsacrum. Two additional follicles are recognized displaced proximally near the left wing. All the follicles are nearly circular, with slight variation in size, ranging from 5.8 to 8.8 mm in diameter (Supplementary Table 1). The follicles are preserved as black carbonization in the slab (Fig. 2a, c) and as impressions in the counterslab (Supplementary Fig. 4). The surfaces of the follicles preserve an uneven lattice structure of fibres, which may potentially be imprints of the blood vessels from the highly vascularized perifollicular membrane⁷.

Enantiornithes indet. STM10-45 (Fig. 2b, d) preserves fewer follicles (at least five), subequal in size (ranging from 6.7 to 8.8 mm in diameter; Supplementary Table 1). The follicles are located on the left of the vertebral column distal to the sternum but cranial to the pelvis; a few disarticulated rib elements overlap the follicles. Four well-preserved follicles are associated together (possibly obscuring more follicles); an additional follicle(s) is caudally displaced slightly from the others (estimated total of five to seven preserved). The surfaces of the follicles are marked by small, black circular spots of mineralization; the roundness of these structures suggests that preservation of the follicles in STM10-45 may have been bacterially mediated.

The circular structures interpreted as follicles are clearly located in the body cavity of these fossil birds; in both the slab and counterslab they are preserved overlapping with bone. The structures are not considered to be seeds preserved in the stomach on the basis of their morphology and position (E. M. Friis, personal communication), as well as comparison with other types of seeds observed in Jehol fossils. In *Jeholornis prima* IVPP (Institute of Vertebrate Palaeontology and Paleoanthropology, Beijing) V13274, which preserves seeds in its stomach¹⁰, the seeds are more numerous, larger (about 10 mm in diameter), differ in shape (tapered at one end, not circular) and the surface is ornamented. The seeds are distributed more ventrally whereas the follicles in *Jeholornis* STM2-51 are dorsally located along the vertebral column and just cranial to the pelvis, consistent with the position of the ovaries in living archosaurs. Seeds are also preserved in the crop of some Jehol birds (*Sapeornis*, *Hongshanornis*)¹¹; in these

¹Institute of Geology and Paleontology, Linyi University, Linyi, Shandong 276000, China. ²Tianyu Natural History Museum of Shandong, Pingyi, Shandong 273300, China. ³Key Laboratory of Vertebrate Evolution and Human Origin of the Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. ⁴Department of Paraclinical Sciences, Faculty of Veterinary Science, University of Pretoria, Private Bag X 04, Onderstepoort 0110, South Africa.

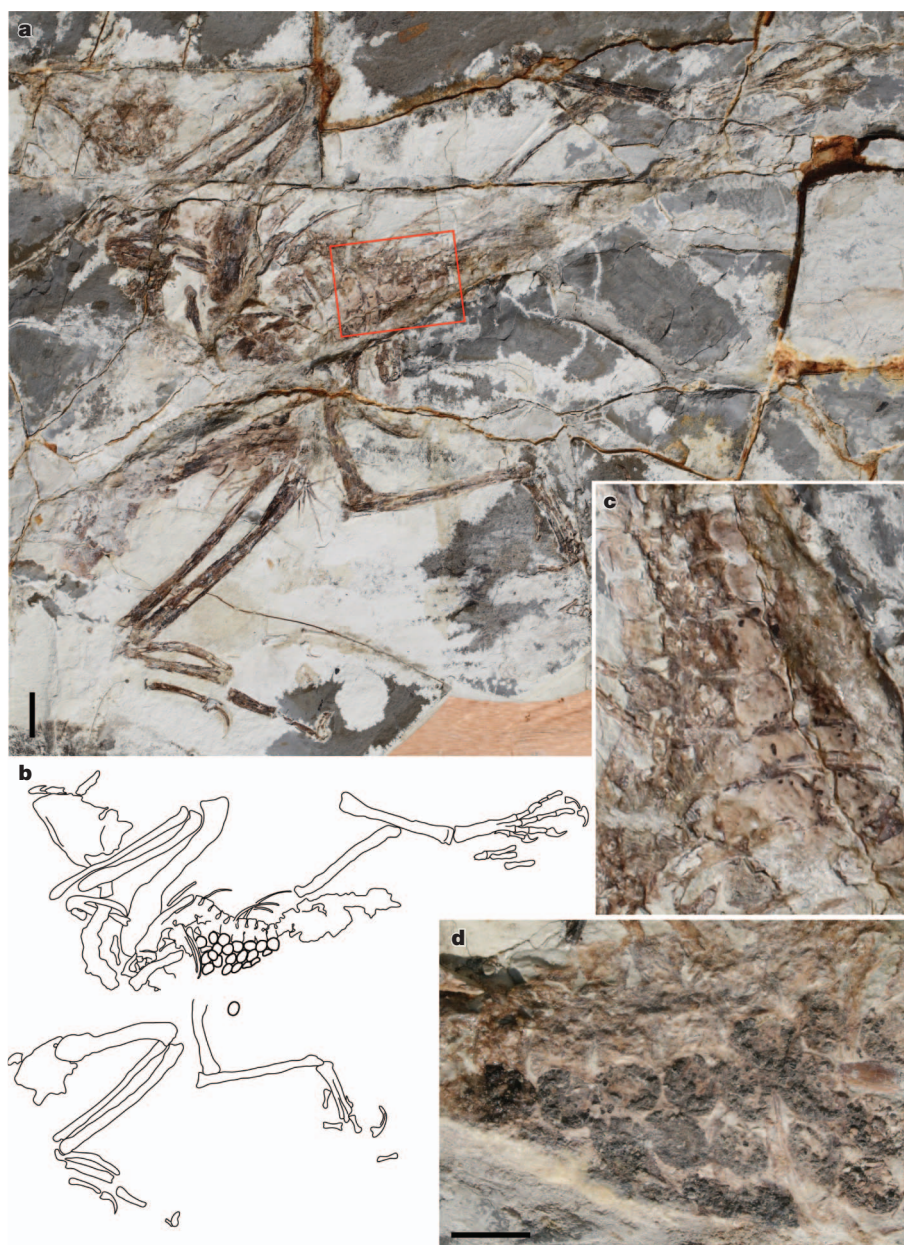


Figure 1 | *Jeholornis* sp. STM2-51 preserving mature ovarian follicles.

a, Photograph of the main slab in dorsal view. Scale bar, 2 cm. **b**, Interpretative drawing of the slab. **c**, Close-up of the follicles indicated by the red square in the

main slab photograph. Scale bar, 1 cm. **d**, Close-up of the follicles in the counterslab. Scale bar, 1 cm.

specimens, the crop is found cranial to the sternum and the seeds differ from the follicles in that they are uneven in size and morphology (less round). These structures are clearly not gastroliths, which are commonly preserved in a large number of Jehol ornithuromorph birds¹¹; these stones are always preserved nearly as in life: three-dimensional, numerous, varying in shape, size and mineral composition, with polished surfaces. Furthermore, gastroliths are proportionately much smaller than the soft-tissue structures preserved here. The circular shape of the preserved structures is consistent with two-dimensional preservation of a spherical structure, such as the mature follicles of the ovary (single enlarged cells). The absence of any preserved eggshell and their clumped association, with no caudally located *in situ* follicles, indicates that the follicles had not entered the oviducts.

Living archosaurs—crocodilians and birds—differ greatly in their reproductive habits in terms of clutch size, nesting behaviour, degree of parental care and developmental strategy of young, posing the evolutionary questions of when, how and why derived avian reproductive

traits evolved within Dinosauria. The evidence indicates that although the non-avian theropod reproductive system retained some primitive archosaurian traits (that is, two functional oviducts, hyper-ellipsoidal eggs, precocial offspring), derived bird-like characteristics were also present (one egg per oviduct per day deposited, larger eggs relative to body size, complex eggshell microstructures, asymmetrical eggs, parental care)^{1,5,6,12}. However, information on the soft-tissue anatomy of reproductive organs and the sequence and timing of changes in reproduction strategy within the Aves class itself remains poorly understood. These new specimens provide the first glimpse into how basal birds reproduced. They also record a fairly wide bracket within the Mesozoic bird phylogenetic tree (Fig. 3), thus revealing different stages of evolution.

In living archosaurs, as in all vertebrates, the female reproductive system is generally divided into two separate parts: the ovary and oviduct^{13,14}. Birds are unique among amniotes in that, although the embryo has two ovaries and oviducts, only the left ovary and oviduct

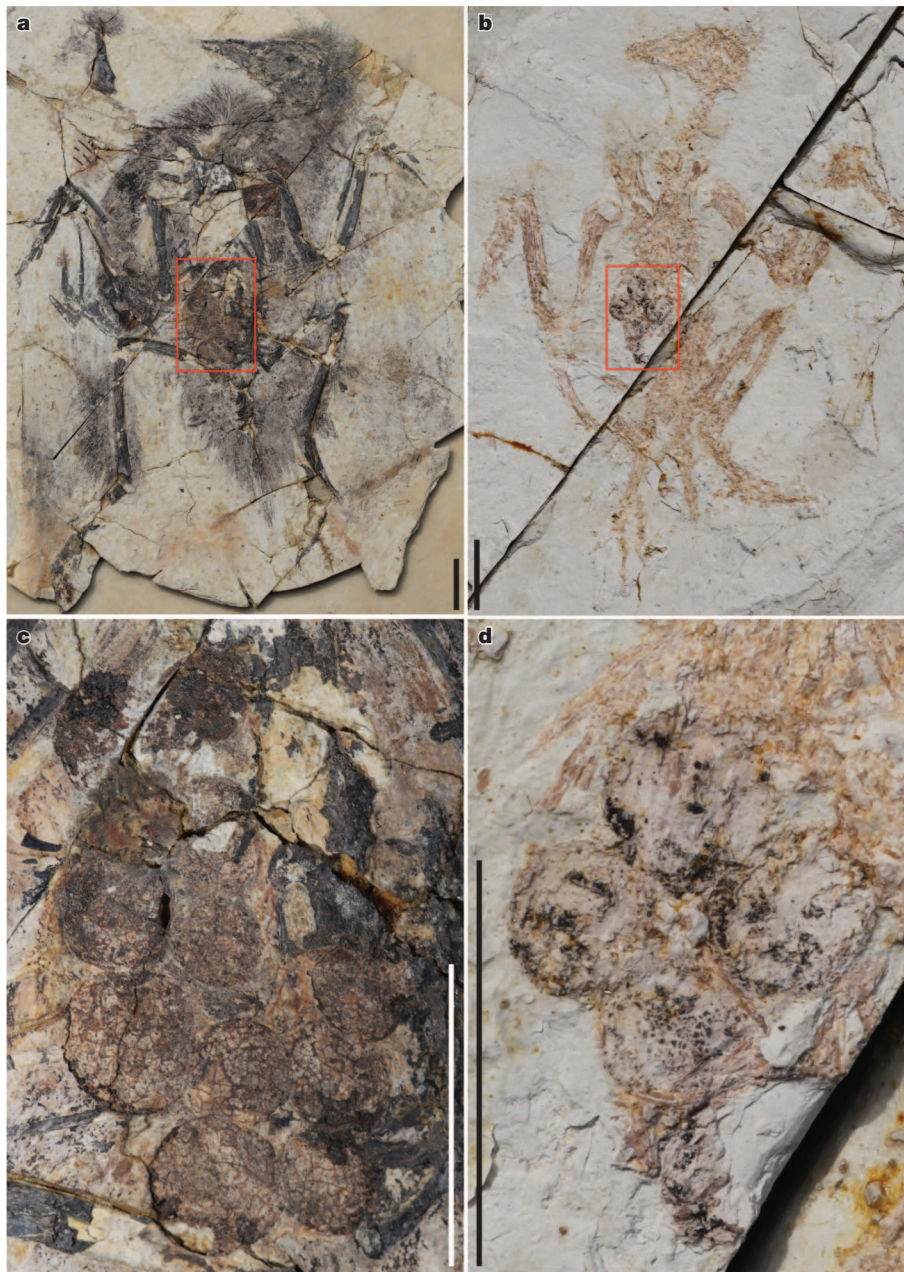


Figure 2 | Enantiornithine specimens preserving mature ovarian follicles in dorsal view.

a, Photograph of the main slab of STM29-8. **b**, Photograph of the main slab of STM10-45. **c**, Close-up of the follicles in STM29-8 preserved in the region indicated by the red box in **a**. **d**, Close-up of the follicles in STM10-45 preserved in the region indicated by the red box in **b**. All scale bars, 2 cm.

are functional in the adult. The kiwi differs from other birds in that both the left and right ovaries develop (and are functional), although it is only the left oviduct that develops (other exceptions occur, particularly among birds of prey)¹⁴. The ovary contains all of the female reproductive oocytes, the number of which diminishes during the lifespan of the individual, although the ovary itself enlarges as the bird approaches reproductive maturity. Each reproductive cycle some of the normally small oocytes mature into follicles, becoming enlarged through yolk deposition (vitellogenesis) in preparation for ovulation. In crocodiles, similar to birds, the follicles do not mature all at once, forming a hierarchy that reflects the order in which follicles will ovulate. Crocodilians have a low metabolic rate and therefore vitellogenesis lasts an extended period of time (months compared to 4–16 days in birds)^{14,15}, as a result, the follicular hierarchy is much less distinct in crocodilians and mature follicles are subequal in size¹⁵. Living birds have a high metabolic rate and yolk deposition is rapid, resulting in strongly apparent follicular hierarchy⁷ (Fig. 3). No medullary bone (a rapidly mobilized source of eggshell calcium in living birds) forms before oviposition^{16,17}; instead, crocodilians access calcium from structural

bone¹⁸. Female crocodilians will ovulate a full clutch of up to 60 eggs (average clutch size ranges from 12 to 48 among living crocodilians)¹⁹, and store the eggs in the distal end of the oviduct for several weeks, with egg laying occurring en masse when the embryos are approximately at the 18-somite stage¹⁸. Although the adult guards the nest and young, eggs are passively incubated and young are not typically fed by the adult, a highly precocial developmental strategy. Living neornithine birds oviposit daily, with the exception of the paleognaths, which lay one egg every 2–6 days¹⁴.

Non-avian maniraptoran dinosaurs are known to have two active oviducts, from the paired association of eggs in the body cavity of one specimen⁵. Why living birds lost the right ovary and oviduct is unclear, but the most common hypothesis suggests that this was related to the need to reduce weight in flight during the reproductive season, the female having to carry only a single egg inside rather than two, although loss may also have been to alleviate the demands for calcium during ovulation²⁰. As the follicles are interpreted as being in the ovaries of these specimens, we cannot comment directly on the number of functioning oviducts; however, both enantiornithines, in dorsal

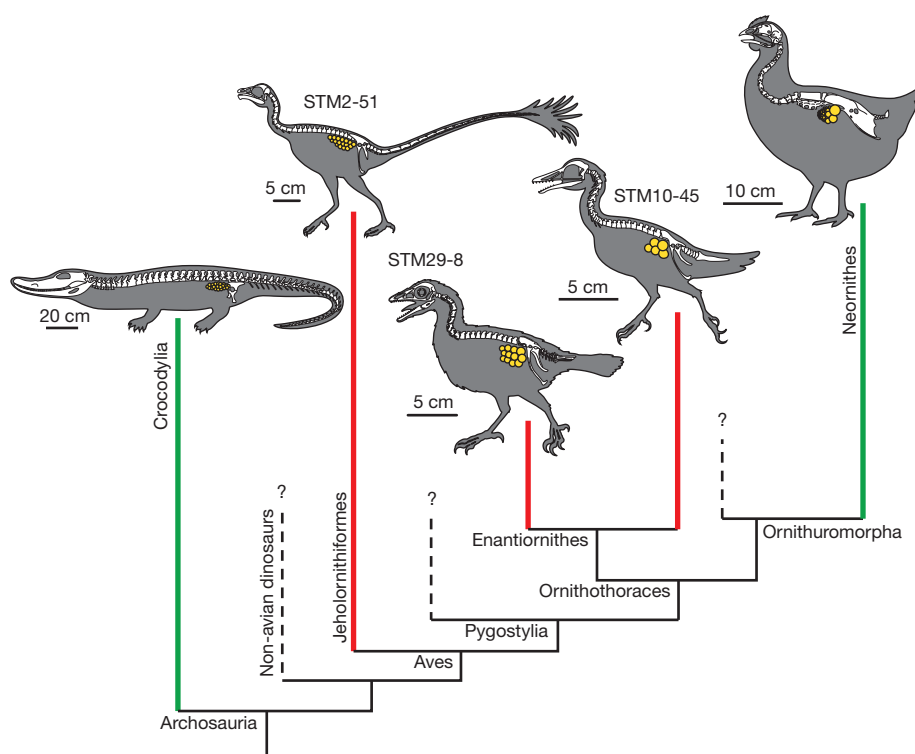


Figure 3 | Reconstructions of living archosaurs *Crocodylia* and *Neornithes*, as well as the three fossilized specimens showing relative ovary morphology mapped over a simplified cladogram of Archosaur relationships. Green lines indicate extant lineages, red lines indicate extinct lineages and dashed lines indicate lineages with no data.

view, preserve the association of follicles clearly on the left side of the body. The position of the ovary is not clear in *Jeholornis* STM2-51; however, the thoracic vertebrae dorsal to the follicles are preserved in left lateral view and the follicles appear to be isolated to the left side, as in the enantiornithines, suggesting that only the left ovary was active. As the closest-related dinosaurs are known to have two functioning oviducts, the loss of function in the right ovary is believed to have occurred very close to the non-avian maniraptoran–avian transition, supporting the hypothesis that its loss is flight related. Notably, it is documented in some crocodilians that the right ovary is the larger of the two, producing a greater number of mature follicles (and thus having a greater weight)¹⁵.

The large and even size of the follicles preserved in these specimens suggests that these female individuals were ready or near ready to ovulate; if the birds were in the early stages of vitellogenesis, follicle size would not be so large and consistent. The slightly higher variation in size among the follicles preserved in STM29-8 may suggest that vitellogenesis was incomplete. Alternatively, this may indicate that vitellogenesis occurred more rapidly and over a shorter period of time, producing a more distinct follicular hierarchy, similar to living birds. This latter hypothesis is consistent with the smaller number of follicles relative to *Jeholornis* and the more derived phylogenetic placement of enantiornithines. The ovaries are not preserved in their entirety. We infer that the immature oocytes, previtellic follicles and atretic follicles in the ovary were either not preserved or are not visible (small, obscured by overlap), or potentially a combination of both. We propose that the presence of a perivitelline layer and other protective layers (absent in immature oocytes)⁷ facilitated preservation of the mature follicles (see Supplementary Information). The absence of eggs in the oviduct(s) suggests that ovulation had not yet begun; alternatively, the animal may have died in the time between laying one egg and ovulating another (typically less than 1 hour in living birds), or an egg(s) in the oviduct(s) may have been expelled from the body after death, similar to the female pterosaur reported from the Jurassic of China²¹.

In the long-tailed bird *Jeholornis* STM2-51 the size of the follicles is very consistent (Fig. 1c, d), suggesting a more crocodilian style of reproduction, in which a large number of follicles reach maturity near

simultaneously so that follicular hierarchy is minimal²². This suggests that vitellogenesis occurred more slowly, similar to crocodilians, and is consistent with the lower metabolic rate relative to extant taxa predicted for basal birds²³. The preserved size of the follicles suggests that egg size was large relative to the body in basal birds, despite morphological restrictions such as distally contacting pubes; eggs would have been too large to store in the body and so oviposition would have to occur between ovulation, as in other maniraptorans and living birds, but unlike crocodiles⁵.

In a healthy and undisturbed female, all mature follicles will ovulate and thus the number of mature follicles is a good proxy for clutch size^{7,15}. Crocodilians have much larger clutches than birds, but the follicle (and complete egg) is much smaller compared to the adult body size. In *Alligator mississippiensis* the mature follicle has a width of 45 mm; the follicle/femur ratio changes with age, but averages 0.017 (refs 15, 24). Comparative data on follicle size is limited, but in *Gallus* mature follicles reach 40 mm (ref. 7) (size exaggerated through domestication), comparable to that of *A. mississippiensis* despite their vast difference in body size (follicle diameter/femur length ratio 0.51). Although the absolute size of the preserved follicles, inferred to be slightly exaggerated by compression, is similar between the three specimens, the adult body size differs considerably: *Jeholornis* STM2-51 body mass is estimated at 676 g; enantiornithine STM29-8 is estimated to be 125 g; enantiornithine STM10-45 is estimated to be 105 g (ref. 25 was used to estimate body mass; Supplementary Table 2). As a result, each specimen has a distinct egg/body size ratio. *Jeholornis* sp. STM2-51 preserves the largest number of follicles, which are proportionately the smallest compared to overall body size (follicle diameter/femur length ratio 0.087), consistent with the basal position of this taxon within Aves (only more derived than *Archaeopteryx*). The larger enantiornithine STM29-8 preserves a larger number of smaller follicles (follicle/femur ratio 0.171), whereas the smaller enantiornithine STM10-45 preserves only a few follicles nearly the same size as *Jeholornis* STM2-51, despite the disparity in their overall body size (follicle/femur ratio 0.217). These specimens show the same trade-off between egg size and clutch size observed in living birds²⁶. The large clutch size in *Jeholornis* STM2-51 and Enantiornithes indet. STM29-8 is consistent with the precocial reproductive strategy inferred for basal birds²⁷. The

smaller clutch preserved in *Enantiornithes* indet. STM10-45 (although potentially an artefact of preservation) suggests that enantiornithines may also have been diverse in their reproductive strategies.

Another key difference between living archosaurs is the timing of sexual maturation. In crocodilians reproductive maturity has an early onset, before the animal reaches skeletal maturity. Modern birds grow rapidly, typically reaching skeletal maturity within 1 year, although not typically becoming sexually mature until later (2–8 years; only 6 months in the domestic chicken)^{7,14}. Histological analyses of paravian dinosaurs reveal a crocodilian pattern of reaching sexual maturity before skeletal maturity²⁸. Enantiornithine STM10-45 shows a clear absence of fusion in compound bones like the carpometacarpus (Supplementary Fig. 8). In *Enantiornithes*, compound bones are known to fuse late in ontogeny⁸ and STM10-45 is regarded as a subadult, indicating that, in at least this lineage of enantiornithines, reproductive maturity was achieved before skeletal maturity, as in paravian dinosaurs and crocodilians. Histological analysis was attempted on the three specimens, but unfortunately no bone could be sampled from STM10-45. Samples from STM2-51 and STM29-8 were badly crushed and the medullary cavity was not clearly preserved, preventing the identification of medullary bone (see Supplementary Information). However, it was determined that enantiornithine STM29-8, which shows a greater degree of fusion in the compound bones (poor preservation prevents determining the exact degree of fusion), was skeletally mature and required more than 1 year to reach adult size (see Supplementary Information).

The preservation of reproductive organs such as ovarian follicles allows a rare opportunity to confidently identify the gender of these specimens and recognize sexually dimorphic traits. The elongate paired tail feathers present in some enantiornithines strongly resemble those of *Confuciusornis*, which have been interpreted as a sexually dimorphic trait present only in males; this interpretation is controversial as the feathers do not co-vary with size as expected²⁹. Feathers are only preserved in STM29-8 (Fig. 2a). The feathers around the pygostyle are clearly preserved and confirm the absence of paired elongate rectrices in this reproductively mature (adult) female.

The ovarian follicles preserved in these Early Cretaceous birds, the first discovery of this kind, reveal new clues regarding basal bird reproduction. Based on the bracket placed by the fossil record, the consistent preservation of the follicles on the left side of the body in all three specimens suggests that the right functional ovary was lost at the dinosaur–avian transition. Clutch size in basal birds is inferred to be plesiomorphically large, consistent with the precocial developmental strategy inferred from embryos and juveniles, and information from closely related dinosaurs. Within Aves there is an observable trend towards smaller clutches with relatively larger eggs that extends from long bony-tailed birds closely related to *Archaeopteryx* into Neornithes, and clade-specific specializations in reproduction strategy were not limited to the crown group. Despite the presence of advanced features, the ovary in Mesozoic birds was large with minimal follicular hierarchy, and sexual maturity preceded skeletal maturation as in crocodilians, both conditions consistent with the lower metabolic rate inferred for basal birds and paravian dinosaurs compared to neornithine birds.

METHODS SUMMARY

Bone samples were taken from all three specimens; only a single sample could be retrieved from STM10-45 and it could not be used for histological analysis. Two samples were taken from each of the other specimens where the bone was already broken, as close to midshaft as preservation allowed: the ulna and femur were sampled in STM2-51 and the humerus and femur were sampled in STM29-8. The samples were taken using a micro-saw and were embedded in EXAKT Technovit 7200 one-component resin and allowed to dry for 24 h. The samples were then cut and polished until the desired optical contrast was reached. The samples were viewed under normal and polarized light using a Leica DMRX polarizing microscope.

Received 2 November 2012; accepted 5 February 2013.

Published online 17 March 2013.

1. Zelenitsky, D. in *Proceedings of the 2006 Goseong International Dinosaur Symposium* (ed Lee, Y.-N.) Vol. 22, 209–216 (J. Paleontol. Soc. Korea, 2006).
2. Zhou, Z. & Zhang, F. A precocial avian embryo from the Lower Cretaceous of China. *Science* **306**, 653 (2004).
3. Chiappe, L. M., Ji, S. & Ji, Q. Juvenile birds from the Early Cretaceous of China: implications for enantiornithine ontogeny. *Am. Mus. Novit.* **3594**, 1–46 (2007).
4. Norell, M. A., Clark, J. M., Chiappe, L. M. & Dashzeveg, D. A nesting dinosaur. *Nature* **378**, 774–776 (1995).
5. Sato, T., Chang, Y.-N., Wu, X.-C., Zelenitsky, D. A. & Hsiao, Y.-F. A pair of shelled eggs inside a female dinosaur. *Science* **308**, 375 (2005).
6. Varricchio, D. J., Jackson, F., Borkowski, J. J. & Horner, J. R. Nest and egg clutches of the dinosaur *Troodon formosus* and the evolution of avian reproductive traits. *Nature* **385**, 247–250 (1997).
7. Gilbert, A. B. in *Form and Function in Birds* (eds King, A. S. & McLelland, J.) 237–360 (Academic, 1979).
8. O'Connor, J. K. A *Systematic Review of Enantiornithes (Aves: Ornithothoraces)*. PhD thesis, Univ. Southern California (2009). 600 pp.
9. O'Connor, J. K., Chiappe, L. M. & Bell, A. in *Living Dinosaurs: the Evolutionary History of Birds* (eds Dyke, G. D. & Kaiser, G.) 39–114 (J. Wiley & Sons, 2011).
10. Zhou, Z. & Zhang, F. A long-tailed, seed-eating bird from the Early Cretaceous of China. *Nature* **418**, 405–409 (2002).
11. Zheng, X. *et al.* Fossil evidence of avian crops from the Early Cretaceous of China. *Proc. Natl Acad. Sci. USA* **108**, 15904–15907 (2011).
12. Clark, J. M., Norell, M. A. & Chiappe, L. M. An oviraptorid skeleton from the Late Cretaceous of Ukhaa Tolgod, Mongolia, preserved in an avianlike brooding position over an oviraptorid nest. *Am. Mus. Novit.* **3265**, 1–36 (1999).
13. Girling, J. E. The reptilian oviduct: a review of structure and function and directions for future research. *J. Exp. Zool.* **293**, 141–170 (2002).
14. Gill, F. B. *Ornithology* 3rd edn 399–430 (W. H. Freeman, 2007).
15. Lance, V. A. Reproductive cycle of the American Alligator. *Am. Zool.* **29**, 999–1018 (1989).
16. Schweitzer, M. H., Wittmeyer, J. L. & Horner, J. R. Gender-specific reproductive tissue in ratites and *Tyrannosaurus rex*. *Science* **308**, 1456–1460 (2005).
17. Schweitzer, M. H., Elsey, R. M., Dacke, C. G., Horner, J. R. & Lamm, E.-T. Do egg-laying crocodilian (*Alligator mississippiensis*) archosaurs form medullary bone? *Bone* **40**, 1152–1158 (2007).
18. Richardson, K. C., Webb, G. J. W. & Manolis, S. C. *Crocodyles: Inside Out* 1st edn, 96–104 (Surrey Beatty & Sons, 2002).
19. Thorbjarnarson, J. B. Reproductive characteristics of the order Crocodylia. *Herpetologica* **52**, 8–24 (1996).
20. Taylor, T. G. How an eggshell is made. *Sci. Am.* **222**, 88–95 (1970).
21. Lü, J.-C. *et al.* An egg-adult association, gender, and reproduction in pterosaurs. *Science* **331**, 321–324 (2011).
22. Astheimer, L. B., Manolis, S. C. & Grau, C. R. Egg formation in crocodiles: avian affinities in yolk deposition. *Copeia* **1989**, 221–224 (1989).
23. Erickson, G. M. *et al.* Was dinosaurian physiology inherited by birds? Reconciling slow growth in *Archaeopteryx*. *PLoS ONE* **4**, e7390 (2009).
24. Farlow, J. O., Hurlbert, G. R., Elsey, R. M., Britton, A. R. C. & Langston, W. Jr. Femoral dimensions and body size of *Alligator mississippiensis*: estimating the size of extinct mesoeucrocodylians. *J. Vertebr. Paleontol.* **25**, 354–369 (2005).
25. Liu, D., Zhou, Z.-H. & Zhang, Y.-G. Mass estimate and evolutionary trend in Chinese Mesozoic fossil birds. *Vertebr. Palasiat.* **50**, 39–52 (2012).
26. Godfray, H. C. J., Partridge, L. & Harvey, P. H. Clutch size. *Annu. Rev. Ecol. Syst.* **22**, 409–429 (1991).
27. Varricchio, D. J. *et al.* Avian paternal care had dinosaur origin. *Science* **322**, 1826–1828 (2008).
28. Erickson, G. M., Curry Rogers, K., Varricchio, D. J., Norell, M. A. & Xu, X. Growth patterns in brooding dinosaurs reveals the timing of sexual maturity in non-avian dinosaurs and genesis of the avian condition. *Biol. Lett.* **3**, 558–561 (2007).
29. Chiappe, L. M., Marugán-Lobón, J., Ji, S. & Zhou, Z. Life history of a basal bird: morphometrics of the Early Cretaceous *Confuciusornis*. *Biol. Lett.* **4**, 719–723 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements The specimens reported here are deposited in the Tianyu Natural History Museum, Shandong, China. We thank A. Shi for helping with illustrations, S. Zhang for help with the histological analysis and E. O'Connor for estimating body mass. The research was supported by the National Basic Research Program of China (973 Program, 2012CB821906), the National Natural Science Foundation of China (41172020) and the Chinese Academy of Sciences.

Author Contributions X.Z., J.O'C. and Z.Z. designed the project. J.O'C., F.H., Z.Z., X.W., M.W. and Y.W. performed the research. J.O'C., F.H. and Z.Z. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J. O'C. (jingmai.oconnor@gmail.com) or Z.Z. (zhouzhonghe@ivpp.ac.cn).

Accelerated gene evolution through replication–transcription conflicts

Sandip Paul¹, Samuel Million-Weaver¹, Sujay Chattopadhyay¹, Evgeni Sokurenko¹ & Houra Merrikh¹

Several mechanisms that increase the rate of mutagenesis across the entire genome have been identified; however, how the rate of evolution might be promoted in individual genes is unclear. Most genes in bacteria are encoded on the leading strand of replication^{1–4}. This presumably avoids the potentially detrimental head-on collisions that occur between the replication and transcription machineries when genes are encoded on the lagging strand^{1–4}. Here we identify the ubiquitous (core) genes in *Bacillus subtilis* and determine that 17% of them are on the lagging strand. We find a higher rate of point mutations in the core genes on the lagging strand compared with those on the leading strand, with this difference being primarily in the amino-acid-changing (nonsynonymous) mutations. We determine that, overall, the genes under strong negative selection against amino-acid-changing mutations tend to be on the leading strand, co-oriented with replication. In contrast, on the basis of the rate of convergent mutations, genes under positive selection for amino-acid-changing mutations are more commonly found on the lagging strand, indicating faster adaptive evolution in many genes in the head-on orientation. Increased gene length and gene expression amounts are positively correlated with the rate of accumulation of nonsynonymous mutations in the head-on genes, suggesting that the conflict between replication and transcription could be a driving force behind these mutations. Indeed, using reversion assays, we show that the difference in the rate of mutagenesis of genes in the two orientations is transcription dependent. Altogether, our findings indicate that head-on replication–transcription conflicts are more mutagenic than co-directional conflicts and that these encounters can significantly increase adaptive structural variation in the coded proteins. We propose that bacteria, and potentially other organisms, promote faster evolution of specific genes through orientation-dependent encounters between DNA replication and transcription.

Concurrent DNA replication and transcription leads to conflicts that stall replication, especially on the lagging strand, where the two machineries meet head-on (Supplementary Fig. 1)^{5–8}. Presumably, to avoid these encounters that can delay replication, bacteria have co-oriented most of their genes with replication by encoding them on the leading strand^{1,2}. However, it is unclear why, even in species with a strong orientation bias such as *B. subtilis*, 25% of all genes and 6% of all essential genes remain on the lagging strand^{1,2}.

We investigated the relation between gene orientation and mutagenesis, by analysing the rate of mutations of ubiquitous (core) genes from five clonally divergent strains of *B. subtilis* (Supplementary Table 1). To be defined as a core gene, the gene had to be present in all five strains, have at least 95% nucleotide identity and at least 95% of the length coverage, to exclude the highly diverse genes affected by non-homologous gene shuffling⁹. Among the 759 core genes, 132 (17%) were encoded on the lagging strand (Supplementary Table 2). Of the 148 core genes that were determined to be essential¹⁰, only six (4%) were on the lagging strand, consistent with the previously described strong orientation bias in the essential genes^{1,2}.

We compared the rates of silent or synonymous (dS), and amino-acid-changing or nonsynonymous (dN) mutations, of the core genes on the leading and lagging strands. The rate of synonymous mutation on the lagging was marginally (2%) higher than on the leading strand (Fig. 1a). In contrast, the rate of nonsynonymous mutations of the lagging strand genes was 42% higher than the genes on the leading strand (Fig. 1b, $P < 0.0001$), without a significant number of outliers (Supplementary Fig. 2). The relative increase of dN in the genes on the lagging compared with the leading strand was irrespective of the dS amount of the corresponding genes (Supplementary Fig. 3). We found a similar trend when we analysed the mutagenesis rate in all leading- or lagging-strand genes: that is, besides the core genes, on the two strands (Supplementary Table 3). However, among the core genes that were previously identified as essential, we did not detect a difference in structural variability between the two strands (Supplementary Fig. 4).

The observed difference between the rates at which leading- and lagging-strand genes vary may be due to orientation-dependent encounters between replication and transcription. It was previously suggested that transcript length contributes to severity of conflicts on the basis of the observation that gene operons oriented head-on tend to be of a relatively small size⁴. Interestingly, we found that most (82%) core genes on the lagging strand are not grouped together, and are organized as single genes rather than operons (Supplementary Fig. 5). Moreover, there is a significant bias for genes on the lagging strand to be significantly shorter on average, with the average gene size being 48% larger on the leading strand (681 compared with 459 base pairs). In particular, genes coding for proteins with a length longer than 200 amino acids are underrepresented on the lagging strand, with only 26% of genes (34 of 132 genes) being of this size category, whereas 48% of genes exceed 200 amino acids on the leading strand (Fig. 2a, $P < 0.0001$). These data are consistent with selection for genes on the lagging strand to be shorter, potentially to decrease the frequency of head-on conflicts.

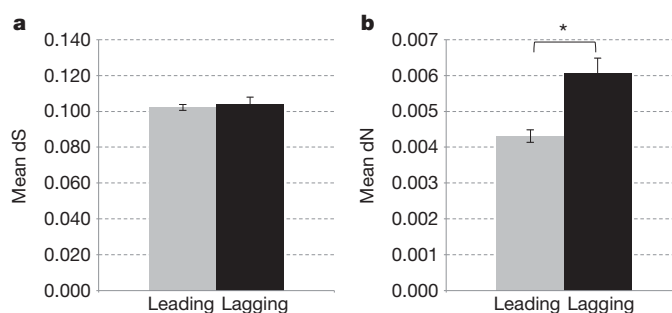


Figure 1 | Highly conserved ‘core’ genes show a higher rate of nonsynonymous mutations on the lagging compared with the leading strand. The mean values for the core genes in relation to diversity are presented. The dS (a) and dN (b) values are plotted for the leading (grey) and lagging (black) strands. Error bars, s.e.m. A statistically significant difference in dN between the two strands is detected with $*P < 0.0001$. Analysis of statistical significance was performed using the Z-test for dS and dN values (see Methods).

¹Department of Microbiology, Health Sciences Building—J-wing, University of Washington, Seattle, Washington 98195, USA.

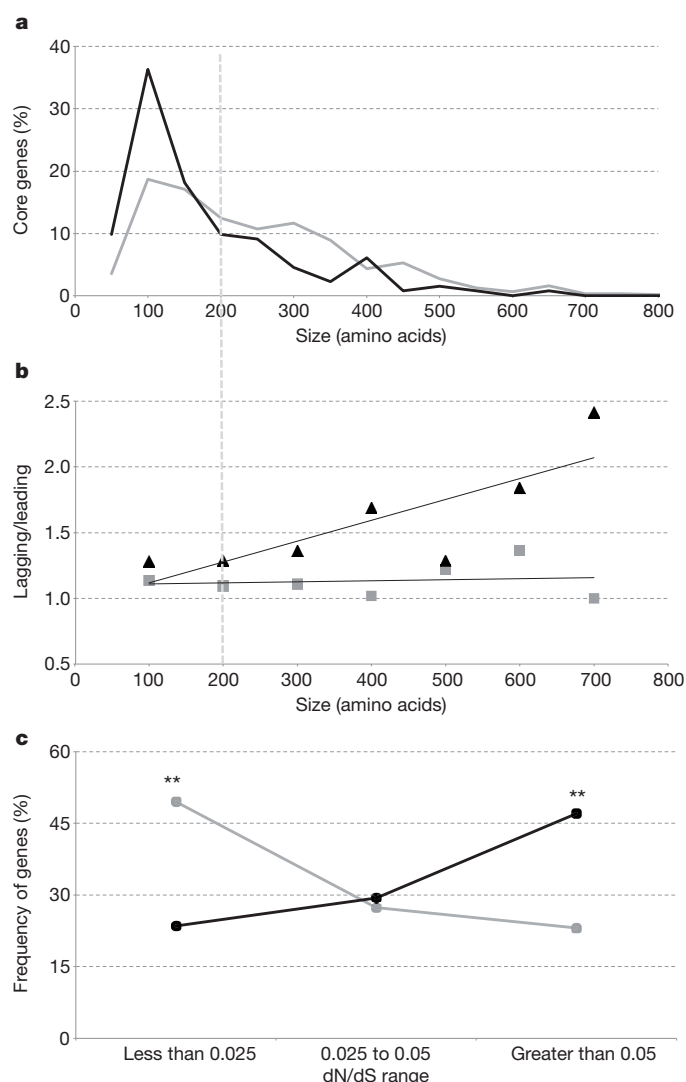


Figure 2 | For head-on genes, a significant positive correlation exists between increased length, mutagenesis and positive selection. **a**, The percentage of genes in each size category for lagging (black) and leading (grey) strands in increasing windows of 50 amino acids. **b**, The fold difference between either dN (black) or dS (grey) for lagging compared with the leading-strand genes in seven size categories ($R^2 = 0.65$ for dN and 0.02 for dS, over length). **c**, dN/dS ratios for genes longer than 200 amino acids on the leading (grey) and lagging (black) strands (** $P < 0.005$). P values were determined using a 2×2 χ^2 test.

Because gene size probably correlates with conflict level, we analysed the relation between nonsynonymous mutation rates and gene length. Indeed, there was a significant positive correlation between increased length and increased dN on the lagging compared with the leading strand (Fig. 2b), with the relative difference in dN being most apparent among the genes coding for proteins longer than 200 amino acids. In this size category (303 genes on the leading strand and 34 genes on the lagging strand), we also find a positive correlation between transcript abundance and dN, with the mean expression values being threefold higher in the group with higher dN amounts compared with those with the lower dN amounts, only in the lagging-strand genes (Supplementary Fig. 6 and Supplementary Table 5). Together, the positive correlation between increased mutation rates and gene length, as well as expression levels, suggest that replication–transcription conflicts are probably responsible for the increased mutation rates in the lagging-strand genes.

We examined the ratio of dN to dS, to evaluate the relative strength of negative selection against structural changes in the genes coding

Table 1 | Distribution of genes with convergent amino-acid mutations in lagging and leading strands

Strand	Number of genes*	Genes with convergent amino-acid mutations	2×2 χ^2 P value
Lagging	34	8	0.04
Leading	303	34	

* Genes that encode proteins with length of 200 or more amino acids.

proteins more than 200 amino acids in length. A significantly higher proportion of the genes on the leading strand (about half) are in the very low dN/dS range of less than 0.025 compared with those on the lagging strand ($P = 0.004$, Fig. 2c). Furthermore, there are 25 genes on the leading strand but none on the lagging strand that lack any structural variation at all. Thus, genes that are under especially strong selection against nonsynonymous changes are heavily biased to be co-oriented with replication. In contrast, almost half of the genes on the lagging strand had a relatively higher dN/dS range of more than 0.05, significantly more than that on the leading strand ($P = 0.002$). To determine whether the increased dN/dS on the lagging-strand genes could be in part due to selection for nonsynonymous mutations, we looked for the presence of convergent amino-acid mutations, which is strong evidence for positive selection and adaptive evolution^{11–13}. Among the genes encoding proteins longer than 200 amino acids, convergent mutations were detected in 24% of the core genes on the lagging strand in contrast to only 11% of the genes on the leading strand ($P < 0.04$) (Table 1 and Supplementary Table 4), indicating faster adaptive evolution of the corresponding genes.

We next determined whether replication–transcription conflicts are responsible for the increased mutation rates in the lagging-strand genes experimentally, using classic reversion assays. The positive correlation between increased gene length, as well as expression levels with increased dN in the lagging-strand genes, strongly suggest that conflicts are driving the mutagenesis of the head-on genes (Fig. 2b and Supplementary Fig. 6). We engineered strains auxotrophic for histidine biosynthesis, and integrated an ectopic copy of the histidine synthetase (*hisC*) gene with a premature stop codon at the 318th position, under the control of the isopropyl-D-thiogalactopyranoside (IPTG) inducible promoter $P_{\text{spank(hy)}}$ onto the chromosome, in either the head-on or co-directional orientation (Fig. 3). Reversion assays and

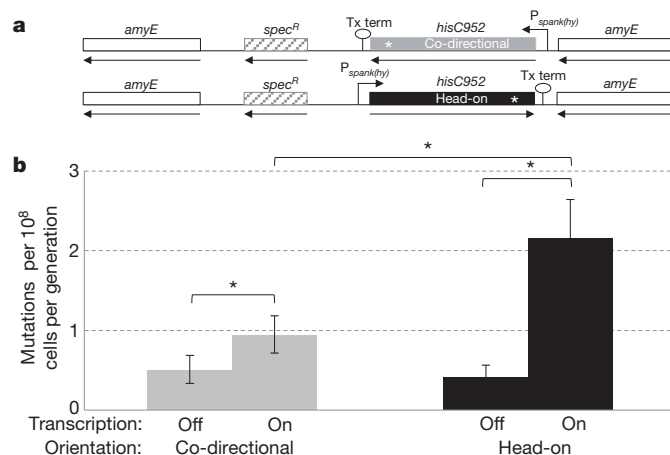


Figure 3 | Increased rates of mutagenesis in the head-on orientation are transcription dependent. **a**, The *amyE* locus containing reporter co-directionally (HM420) or head-on (HM419) to replication. **b**, Mutation rates on the basis of reversion assays for *hisC952* when the gene was oriented head-on or co-directionally to replication, in the presence (transcription on) or absence (transcription off) of IPTG (* $P < 0.05$). Error bars, 95% confidence intervals. We confirmed by sequencing that the transcription-dependent mutations were indeed occurring in the *hisC* allele at *amyE* (Supplementary Table 9). YB955 did not show an increase in *hisC952* reversions with IPTG treatment (data not shown).

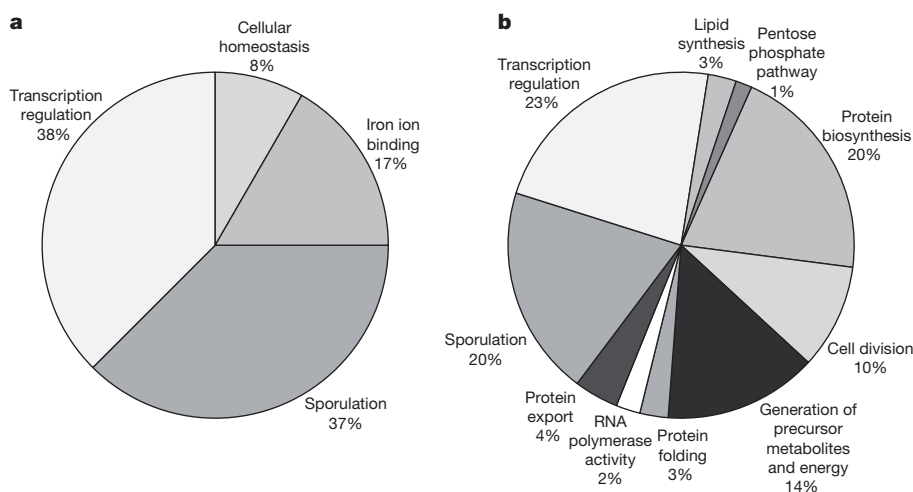


Figure 4 | Functional categories of leading- and lagging-strand genes. Pie charts presenting significantly ($P < 0.05$) overrepresented functional categories (as determined by medium stringency using DAVID²³) of protein products from lagging (a) and leading (b) strands of replication for *B. subtilis* strain 168.

fluctuation analyses were performed in the presence and absence of transcription, for both orientations (see Methods). In the absence of transcription, the orientation of the gene did not affect mutation rates (Fig. 3 and Supplementary Table 6). When activated, transcription led to an increase in the mutation rate in both orientations, but the increase was much more prominent for the head-on orientation (Fig. 3 and Supplementary Table 6). The dependence of increased mutation rates on both transcription and orientation indicates that head-on replication–transcription conflicts lead to increased mutagenesis. These data demonstrate that, in the absence of transcription, mutagenesis rates of the two strands are not significantly different.

We analysed the functional categories of the core genes on the lagging strand and found enrichment in genes for (1) sporulation, (2) iron binding, (3) transcription regulation and (4) cellular homeostasis (Fig. 4). In general, most of these genes represent a variety of stress responses. For example, the transcription regulation genes include extracytoplasmic function (ECF)-type sigma factors (SigV and SigM), which are known to be the most variable of the sigma factors and are involved in responding to extracytoplasmic functions, as well as master regulators of biofilm formation (SinI and SinR), which control the transition from a free-swimming lifestyle to a stress-tolerant, biofilm lifestyle. We propose that these particular functions are kept relatively variable, possibly to produce population heterogeneity that can more easily respond to rapid changes in the environment.

Taken together, our data indicate that genes positioned on the lagging strand evolve at a significantly higher rate than those encoded on the leading strand, and that the mutagenic nature of the lagging strand is due, at least in part, to the conflicts between replication and transcription. The increased rate of mutagenesis in the genes on the lagging strand is probably detrimental for many, especially essential genes, and is probably responsible for the strong bias for these genes to be co-oriented with replication. Two models were proposed previously to explain the underlying cause of the bias against head-on orientation of genes: (1) the detrimental effect of increased mutagenesis, by Mirkin and Mirkin¹⁴; (2) increased rate of gene transcript truncations, by Rocha and Danchin¹. Experimental work from *B. subtilis*, *Escherichia coli* and *Saccharomyces cerevisiae* has shown that genomic instability (which may increase the rate of mutations) is a potential consequence of head-on conflicts^{8,15,16}. Indeed, our findings show that the bias for the co-directional orientation of genes in bacteria is at least partly determined by negative selection against increased rate of nonsynonymous mutations, consistent with the Mirkin and Mirkin model.

The genes categorized for the lagging strand were *abrB*, *acuA*, *acuB*, *ahpC*, *ahpF*, *antE*, *cotM*, *cspD*, *cueR*, *def*, *fer*, *gerPF*, *hpr*, *ispG*, *katA*, *lexA*, *mrgA*, *nasE*, *sacY*, *sda*, *sigM*, *sigV*, *sinI*, *sinR*, *spoIISB*, *sspB*, *sspD*, *sspl*, *sspK*, *sspl*, *sspM*, *sspO*, *tnrA*, *ycnC*, *ydbP*, *ydgJ*, *yhdK*, *yjbI*, *ynzD*, *ypoP*, *ytzE*, *yutI*, *yuxN* and *ywoH*.

The other side of the coin, however, in contrast to the essential genes, is that the increased mutation rate could also be the reason for the head-on orientation of certain core genes: that is, head-on orientation could be promoted by positive selection. The action of positive selection is apparent from the high rate of convergent evolution in the proteins coded by genes on the lagging strand—a strong indication of their adaptive significance. In contrast to the nonsynonymous changes, synonymous mutations are generally (but not always) considered to be fitness-neutral in nature and their accumulation over time is expected to be driven largely by genetic drift. Thus accumulation of synonymous mutations could be a relatively slow process compared with positive selection for the adaptive changes, possibly explaining the low level of accumulation of the synonymous mutations in the lagging-strand genes compared with the nonsynonymous mutations. Although it is possible that there is a stronger negative selection against synonymous mutations in the lagging-strand genes owing to codon usage bias in highly expressed genes, we did not see a large difference between codon usage in the two strands' genes (Supplementary Fig. 7).

It is unclear how the phenomenon, described here, extends to other Gram-positive bacteria or to Gram-negative organisms such as *E. coli* and *Salmonella*. Orientation, transcription and rates of evolution in core genes have not been systematically investigated in these organisms, and previous studies looking at some aspects of these questions have produced contradictory results depending on methodology^{17–22}.

Replication–transcription conflict-mediated mutagenesis could be used by many organisms, including eukaryotes, as a universal strategy to link gene expression and evolution rate under selection. Genes on the lagging strand can evolve at a relatively fast rate because of head-on conflicts. Thus a simple switch in orientation, using short sequence homologies and recombination-dependent mechanisms, could facilitate evolution in specific genes in a targeted way. Investigating the main targets of conflict-mediated mutagenesis is likely to show far-reaching biological insights into adaptation and evolution of organisms.

METHODS SUMMARY

Strains are listed in Supplementary Tables 1 and 7. Plasmids are listed in Supplementary Table 8. Relevant properties of strains are described in the text. Strain constructions, growth conditions and bioinformatics analyses are described both in Methods and Supplementary Methods. Standard procedures were used for the reversion assays and calculation of mutation rates and are described (Methods). The rate of mutations (dS and dN) were determined on the basis of the possible number of changes compared with the actual number of changes observed, either for synonymous or nonsynonymous mutations (see Methods for details).

Full Methods and any associated references are available in the online version of the paper.

Received 8 November 2012; accepted 7 February 2013.

1. Rocha, E. P. & Danchin, A. Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res.* **31**, 6570–6577 (2003).
2. Rocha, E. P. & Danchin, A. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nature Genet.* **34**, 377–378 (2003).
3. Rocha, E. P. The replication-related organization of bacterial genomes. *Microbiology* **150**, 1609–1627 (2004).
4. Price, M. N., Alm, E. J. & Arkin, A. P. Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. *Nucleic Acids Res.* **33**, 3224–3234 (2005).
5. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication–transcription conflicts in bacteria. *Nature Rev. Microbiol.* **10**, 449–458 (2012).
6. Mirkin, E. V. & Mirkin, S. M. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.* **71**, 13–35 (2007).
7. Pomerantz, R. T. & O'Donnell, M. Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science* **327**, 590–592 (2010).
8. De Septenville, A. L., Duigou, S., Boubakri, H. & Michel, B. Replication fork reversal after replication–transcription collision. *PLoS Genet.* **8**, e1002622 (2012).
9. Chattopadhyay, S. *et al.* High frequency of hotspot mutations in core genes of *Escherichia coli* due to short-term positive selection. *Proc. Natl Acad. Sci. USA* **106**, 12412–12417 (2009).
10. Kobayashi, K. *et al.* Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA* **100**, 4678–4683 (2003).
11. Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
12. Christin, P. A., Weinreich, D. M. & Besnard, G. Causes and evolutionary significance of genetic convergence. *Trends Genet.* **26**, 400–405 (2010).
13. Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).
14. Mirkin, E. V. & Mirkin, S. M. Mechanisms of transcription–replication collisions in bacteria. *Mol. Cell. Biol.* **25**, 888–895 (2005).
15. Srivatsan, A., Tehranchi, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet.* **6**, e1000810 (2010).
16. Kim, N., Abdulovic, A. L., Gealy, R., Lippert, M. J. & Jinks-Robertson, S. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair* **6**, 1285–1296 (2007).
17. Veaute, X. & Fuchs, R. P. Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. *Science* **261**, 598–600 (1993).
18. Fijalkowska, I. J., Jonczyk, P., Tkaczyk, M. M., Bialoskorska, M. & Schaaper, R. M. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc. Natl Acad. Sci. USA* **95**, 10020–10025 (1998).
19. Maliszewska-Tkaczyk, M., Jonczyk, P., Bialoskorska, M., Schaaper, R. M. & Fijalkowska, I. J. SOS mutator activity: unequal mutagenesis on leading and lagging strands. *Proc. Natl Acad. Sci. USA* **97**, 12678–12683 (2000).
20. Szczepanik, D. *et al.* Evolution rates of genes on leading and lagging DNA strands. *J. Mol. Evol.* **52**, 426–433 (2001).
21. Lin, C. H., Lian, C. Y., Hsiung, C. A. & Chen, F. C. Changes in transcriptional orientation are associated with increases in evolutionary rates of enterobacterial genes. *BMC Bioinformatics* **12** (suppl. 9), S19 (2011).
22. Juurik, T. *et al.* Mutation frequency and spectrum of mutations vary at different chromosomal positions of *Pseudomonas putida*. *PLoS ONE* **7**, e48511 10.1371/journal.pone.0048511 (2012).
23. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the work of others that has contributed to our understanding of the topic, which we were unable to cite here. We thank B. Brewer, J. Mougous, F. Fang, M. Parsek and S. Miller for reading the manuscript. We thank C. Merrikh for comments on the manuscript and technical help with sequencing of revertants. We thank E. Robleto for the auxotrophic YB955 strain. H.M. was supported by start-up funds from the Department of Microbiology at the University of Washington. S.M.-W. was supported by the University of Washington Biophysics Training Grant. E.S., S.P. and S.C. were supported by National Institutes of Health (NIH) grants RC4 AI092828 and R01 GM084318.

Author Contributions H.M., E.S., S.P. and S.M.-W. designed experiments, S.P., S.C. and S.M.-W. performed the bioinformatics analysis and the experiments, H.M., E.S., S.P., S.C. and S.M.-W. analysed the data. H.M. and E.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S. (evs@u.washington.edu) or H.M. (merrikh@uw.edu).

METHODS

Strains. For the bioinformatics analysis, fully assembled genomes of five *B. subtilis* strains were downloaded from NCBI GenBank. Those strains were *B. subtilis* subsp. *subtilis* str. 168, *B. subtilis* BSn5, *B. subtilis* subsp. *subtilis* str. RO-NN-1, *B. subtilis* subsp. *spizizenii* TU-B-10 and *B. subtilis* subsp. *spizizenii* str. W23 (Supplementary Table 1).

All experiments were performed using isogenic derivatives of *Bacillus subtilis* strain YB955 (*ile*, *leu*, *his*, *met*) (Supplementary Table 7). To construct a strain with an exogenous copy of histidine synthetase oriented head-on to replication, *hisC952* was first amplified from genomic DNA from YB955 using the primers HM386 (AAA AGT CGA CAA GGA GGT ATA CAT TTG CGT ATC AAA GAA CAT TTA AAA C) and HM396 (AAA GCA TGC TCC TTA TGA ATG GGA CTT ATA AAA TTT CAG CTA AAA TGG). The PCR fragment was digested with SalI and SphI and ligated into the integration vector pDRI11 to generate the plasmid pHM392. This plasmid was introduced by double crossover into YB955 at *amyE* to generate strain HM419 (*leu*C427 *met*B5 *hisC952 amyE*::{P_{spank(hy)}-*hisC952* Head-On *spc*}). To construct a strain with the *hisC952* allele oriented co-directionally with replication, a fragment containing *hisC952* and P_{spank(hy)} was amplified off of pHM392 using primers HM392 (AAA GAA TTC TAA CTC ACA TTA ATT GCG TTG C) and HM393 (AAA GGA TCC TGG CAA GAA CGT TGC TCG AGG). This fragment was digested with EcoRI and BamHI, inverted and ligated into pDRI11 to generate plasmid pHM394. This plasmid was introduced by transformation into strain YB955, generating strain HM420 (*leu*C427 *met*B5 *hisC952 amyE*::{P_{spank(hy)}-*hisC952* Co-Directional *spc*}). See Supplementary Table 8 for a list of the plasmids. Candidate spectinomycin-resistant transformants were screened for disruption of the *amyE* locus by starch agar plate assay. Genomic DNA was isolated from amylase-deficient transformants and the chromosomal region containing *amyE* was amplified using primers HM205 (TCC AAA CTG GAC ACA TGG AA) and HM207 (AAA GAG GCG TAC TGC CTG AA). The resulting PCR fragments were sequenced to confirm integration of *hisC952* in the proper orientation.

Detection of common orthologous genes. To identify the core genes shared between all five strains of *B. subtilis*, we used *B. subtilis* subsp. *subtilis* str. 168 as the reference genome and extracted homologues from the other four strains with nucleotide sequence identity and length coverage values of at least 95%. We excluded the annotated pseudogenes and genes with either internal stop codons or non-ACGT characters.

Identification of leading- and lagging-strand genes. We considered the origin of replication to be positioned between ribosomal protein L34 and chromosomal replication initiation protein (DnaA), whereas the terminus region before replication terminator protein (Rtp) for all five strains. Depending on these locations, we defined the genes present in leading- and lagging strands for each strain. For analysis, we considered those genes that were present in the same strand (either leading or lagging) for all five strains. In total we found 759 core polymorphic genes: 627 in the leading strand and 132 in the lagging strand (Supplementary Table 2).

Molecular evolutionary analysis. Alignment of every gene set was performed by clustalW²⁴. The rates of nonsynonymous (dN) and synonymous (dS) mutations for each gene were computed by the ratio of number of nonsynonymous and synonymous changes to the total number of nonsynonymous and synonymous sites respectively, using the mutation-fraction method²⁵. To assess the statistical significance between any two sample sets, a Z-test was applied to dS and dN values²⁶.

Detection of convergent mutations was performed by Zonal Phylogeny software²⁷. Zonal Phylogeny reconstructs an unrooted protein phylogram on the basis of nucleotide sequences where several alleles differentiated by synonymous changes are collapsed into single protein variant, followed by the mapping of phylogenetically unlinked repeated mutations at specific amino-acid positions

(that is, convergent amino-acid mutations). DnaSPv5 was used to calculate the $2 \times 2 \chi^2$ statistics²⁸.

Media and growth conditions. For all experiments, cells were grown at 37 °C, shaking at 260 r.p.m. Liquid cultures of *Bacillus subtilis* were grown in rich medium (Luria-Bertain (LB)) supplemented with spectinomycin (50 µg ml⁻¹) where appropriate. *E. coli* was grown in LB supplemented with ampicillin at 100 µg ml⁻¹. To determine viabilities, *B. subtilis* cells were plated on solid agar plates containing Spizizen's Minimal Medium (0.2 mg ml⁻¹ ammonium sulphate, 1.4 mg ml⁻¹ monobasic potassium phosphate, 0.6 mg ml⁻¹ dibasic potassium phosphate, 0.1 mg ml⁻¹ sodium citrate dihydrate, 0.02 mg ml⁻¹ magnesium sulphate heptahydrate, 100 µg ml⁻¹ glutamic acid, 5 µg ml⁻¹ glucose), supplemented with isoleucine, methionine, leucine and histidine at 50 µg ml⁻¹. To select for reversion to prototrophy, cells were plated on solid agar plates containing Spizizen's Minimal Medium supplemented with isoleucine, methionine and leucine at 50 µg ml⁻¹.

Reversion assays. Fluctuation assays were performed to determine the mutation rate. Strains HM419 and HM420 were grown overnight in LB supplemented with spectinomycin at 50 µg ml⁻¹. Saturated cultures were diluted 1:10⁴ into 2 ml LB medium and 2 ml LB medium supplemented with 1 mM IPTG. For these experiments, 12 parallel cultures for each strain and condition were grown for 5 h at 37 °C with shaking at 260 r.p.m. When cells reached saturation, 100 µl of culture was withdrawn, diluted 1:10⁶ and spread onto solid agar plates containing Spizizen's Minimal Medium supplemented with all required amino acids to determine the number of viable cells present in the culture at plating. The remainder of the culture was pelleted by centrifugation, re-suspended in Spizizen's Minimal Salts and spread onto solid agar plates lacking histidine. Revertant colonies were counted after 48 h of growth at 37 °C. The experiment was repeated three times on different days.

Confirmation of reversions in the *hisC* gene at *amyE*. To determine whether reversions were occurring in the copy of *hisC* at *amyE*, 28 colonies, 24 head-on and four co-directional, from plates with IPTG were sequenced. The revertant colonies were grown up in selective (histidine- spectinomycin+) media and genomic DNA was prepared. The *amyE* region was amplified by PCR with primers HM205 (TCC AAA CTG GAC ACA TGG AA) and HM207 (AAA GAG GCG TAC TGC CTG AA). The PCR products were sequenced using the primer HM396 (AAA GCA TGC TCC TTA TGA ATG GGA CTT ATA AAA TTT CAG CTA AAA TGG). Changes at position 952 were analysed (Supplementary Table 9).

Calculation of mutation rates. Mutation rates were calculated using the Ma-Sandri-Sarkar maximum-likelihood method^{29,30}, with the aid of the Fluctuation Analysis Calculator. Statistical significance was determined using a standard two-tailed *t*-test.

24. Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497–3500 (2003).
25. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
26. Suzuki, Y. & Gojobori, T. in *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny* 1st edn (eds Salemi, M & Vandamme, A.-M.) 283–311 (Cambridge Univ. Press, 2003).
27. Chattopadhyay, S., Dykhuizen, D. E. & Sokurenko, E. V. ZPS: visualization of recent adaptive evolution of proteins. *BMC Bioinformatics* **8**, 187 (2007).
28. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
29. Sarkar, S., Ma, W. T. & Sandri, G. H. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**, 173–179 (1992).
30. Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. Fluctuation analysis Calculator: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).

A central role for TFIID in the pluripotent transcription circuitry

W. W. M. Pim Pijnappel^{1,2,3,†*}, Daniel Esch^{4*}, Marijke P. A. Baltissen^{1,2}, Guangming Wu⁴, Nikolai Mischerikow^{2,5}, Atze J. Bergsma^{1,†}, Erik van der Wal^{3,4,†}, Dong Wook Han^{4,†}, Hermann vom Bruch⁴, Sören Moritz⁴, Phillip Lijnzaad¹, A. F. Maarten Altelaar^{2,5}, Katrin Sameith¹, Holm Zaehres⁴, Albert J. R. Heck^{2,5}, Frank C. P. Holstege¹, Hans R. Schöler⁴ & H. T. Marc Timmers^{1,2}

Embryonic stem (ES) cells are pluripotent and characterized by open chromatin and high transcription levels, achieved through auto-regulatory and feed-forward transcription factor loops^{1,2}. ES-cell identity is maintained by a core of factors including Oct4 (also known as Pou5f1), Sox2, Klf4, c-Myc (OSKM) and Nanog²⁻⁴, and forced expression of the OSKM factors can reprogram somatic cells into induced pluripotent stem cells (iPSCs) resembling ES cells^{5,6}. These gene-specific factors for RNA-polymerase-II-mediated transcription recruit transcriptional cofactors and chromatin regulators that control access to and activity of the basal transcription machinery on gene promoters^{7,8}. How the basal transcription machinery is involved in setting and maintaining the pluripotent state is unclear. Here we show that knockdown of the transcription factor IID (TFIID) complex affects the pluripotent circuitry in mouse ES cells and inhibits reprogramming of fibroblasts. TFIID subunits and the OSKM factors form a feed-forward loop to induce and maintain a stable transcription state. Notably, transient expression of TFIID subunits greatly enhanced reprogramming. These results show that TFIID is critical for transcription-factor-mediated reprogramming. We anticipate that, by creating plasticity in gene expression programs, transcription complexes such as TFIID assist reprogramming into different cellular states.

The basal transcription factor TFIID is central to transcription initiation by RNA polymerase II (Pol II), and is composed of TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs)^{9,10}. TFIID can recognize both specific core promoter sequences as well as chromatin modifications associated with active promoters¹¹⁻¹³. Interestingly, TFIID subunits and TBP paralogues can act in tissue- and development-specific ways^{9,10,13}. We noted high protein levels of the TAF5 and TAF6 subunits of TFIID in mouse ES cells compared to NIH 3T3 cells or a fibroblast cell line (Fig. 1a). Many TAF messenger RNAs are also expressed at higher levels in ES cells (Supplementary Fig. 2). These TFIID subunits are also more abundant in human ES cells and iPSCs (Fig. 1b). To test the relevance of TFIID in mouse ES cells, we applied lentiviral-mediated knockdown and scored colonies for morphology and alkaline phosphatase staining. Notably, reduced expression of many TAFs (TAF1, 3, 4, 4b, 5, 6, 9, 11, 12 or 13) resulted in a differentiated morphology and reduced alkaline phosphatase staining (Fig. 1c and Supplementary Fig. 3). Two exceptions were TAF7 and 8, which despite efficient knockdown (Supplementary Fig. 4) did not affect ES-cell characteristics. Fluorescence-activated cell sorting (FACS) analysis of propidium-iodide-stained cells did not reveal apoptosis or strong proliferation defects after TAF knockdown (Supplementary Fig. 5). In contrast, reduced expression of *Polr2a* (Fig. 1d and Supplementary Fig. 6) or Pol II inhibition by α -amanitin

(Fig. 1e) caused growth defects but failed to induce differentiation. Thus, high levels of TAFs rather than high Pol II activity are required to maintain ES-cell identity, suggesting that TFIID is directly involved in the pluripotent state.

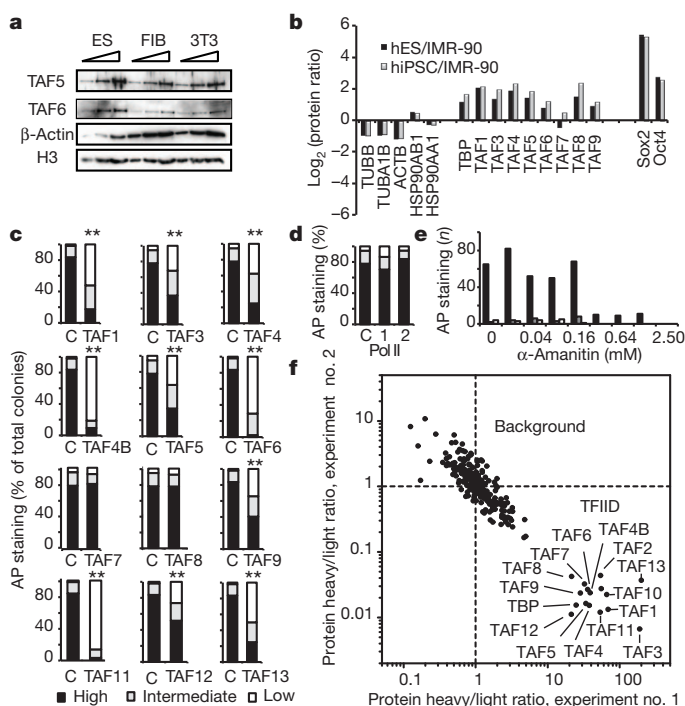


Figure 1 | High expression of TFIID is required for maintaining ES cell identity. **a**, Immunoblot analyses of mouse ES cells (ES), NIH 3T3 cells (3T3), and a stable line of MEFs (FIB) based on equal cell numbers. Twofold dilutions of total cell lysates were loaded as indicated by the histone H3 control. **b**, Quantitative proteomic analyses of human pluripotent cells²⁹. TAFs had an average standard deviation (s.d.) of 40% ($n > 3$). h, human. Values of human (h) iPSCs and ES cells are relative to IMR-90 fetal lung fibroblasts. **c**, Knockdown of TFIID subunits induces ES-cell differentiation. C, control knockdown ($n > 36$). **d**, Knockdown of *Polr2a* (Pol II) fails to induce ES-cell differentiation ($P = 0.18$ and 0.86 , $n > 36$). AP, alkaline phosphatase. **e**, Inhibition of Pol II by α -amanitin does not induce differentiation ($P = 1$ in all cases, $n > 9$). **f**, Quantitative proteomic analysis of TAF6-GFP interactors was performed in duplicate with isotope swap. SILAC ratios are shown as black dots. $^{**}P < 0.01$.

¹Molecular Cancer Research, University Medical Center Utrecht, 3584 CG Utrecht, The Netherlands. ²Netherlands Proteomics Center, 3584 CH Utrecht, The Netherlands. ³Department of Pediatrics and Department of Clinical Genetics, Erasmus Medical Center, 3015 GE Rotterdam, The Netherlands. ⁴Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany. ⁵Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands. [†]Present addresses: Department of Pediatrics and Department of Clinical Genetics, Erasmus Medical Center, 3015 GE Rotterdam, The Netherlands (W.W.M.P.P.; A.J.B.; E.v.d.W.); Department of Stem Cell Biology, School of Medicine, SMART Institute of Advanced Biomedical Science, Konkuk University, 143701 Seoul, Republic of Korea (D.W.H.).

*These authors contributed equally to this work.

TFIID complexes can differ in composition depending on cellular state or cell type^{9,10}. To determine TFIID composition in mouse ES cells, we applied stable isotope labelling by amino acids in cell culture (SILAC)-based quantitative proteomics¹⁴ to green fluorescent protein (GFP)-tagged TAF6 stably expressed at sub-endogenous levels (Supplementary Figs 7 and 8). Isotope ratios of 20 to 200 were observed for TBP and all TFIID TAFs (Fig. 1f and Supplementary Table 1). This shows that TFIID forms the canonical complex in mouse ES cells^{9,10}.

To examine TFIID function further, two independent stable TAF5 or control knockdown ES cells were established. Quantitative polymerase chain reaction with reverse transcription (RT-qPCR) analysis showed reduced mRNA expression of *Nanog*, *Oct4* and *Sox2*, but enhanced expression of the differentiation genes *brachyury* and *Bmp2* in stable TAF5-knockdown cells (Fig. 2a). TAF5-knockdown cells did not show changes in cell proliferation or cell cycle profiles (Supplementary Figs 9 and 10). Genome-wide mRNA profiles showed altered expression of 1,857 genes in TAF5-knockdown cells (using 1.5-fold change with $P < 0.05$; Fig. 2b and Supplementary Fig. 11). Pluripotency genes were downregulated, whereas several neuroectodermal (engrailed 2, neuronatin 1, *Gap43*, *nestin*), mixed ectodermal/

mesodermal (*Pitx2*, *Hoxb8*) and mesodermal (*brachyury*, *Mest*) differentiation markers were upregulated. No known endodermal genes were induced in TAF5-knockdown cells. Gene ontology analysis of downregulated genes in TAF5-knockdown cells showed enrichment for transcription regulators (Supplementary Fig. 12). Affected biological processes included metabolism, development and signal transduction (Supplementary Fig. 13). Interestingly, mRNA profiles of TAF5-knockdown cells showed a high degree of overlap with *Oct4*-knockdown profiles¹⁵ and *Oct4* knockdown reduced expression of several TAFs including TAF4, 6 and 9 (Supplementary Figs 14 and 15). To test the differentiation potential of TAF5-knockdown cells, leukaemia inhibitory factor (LIF) was removed from the cultures and mRNA expression was examined for selected genes. Whereas *brachyury* mRNA was induced to higher levels, *Fgf5* induction was reduced on TAF5 knockdown. Similar to *Oct4*, *Sox2* and *Nanog*, *Taf1* and *Taf5* mRNAs were strongly reduced in ES cells on LIF removal (Supplementary Fig. 16).

These results suggest that high levels of TFIID in ES cells are required to maintain expression of pluripotency but not of differentiation genes. This was tested in chromatin immunoprecipitation (ChIP)-qPCR assays for TAF1 and TBP. Promoters for *Oct4*, *Nanog* and *Klf4* showed reduced occupancy of TAF1 upon TAF5 knockdown (Fig. 2c). In contrast, *Bmp2* and *engrailed-2* promoters showed increased TAF1 binding. ChIP assays for TBP gave similar results, but with a lower recovery. Next, we examined whether TFIID is directly involved in the pluripotent circuitry of ES cells. Published

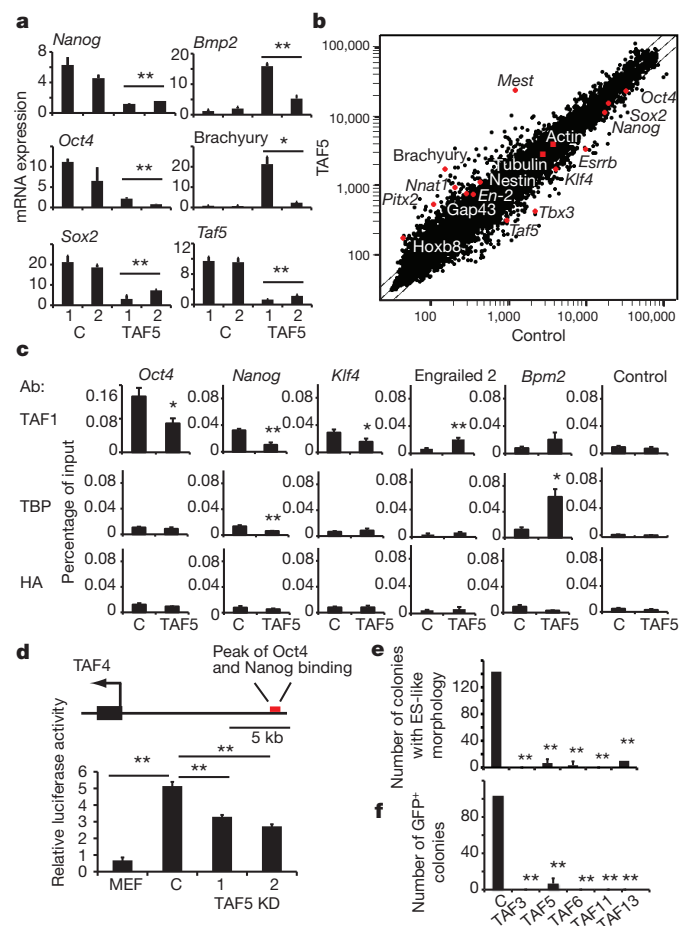


Figure 2 | TFIID is part of the pluripotency circuitry and is required for reprogramming. **a**, RT-qPCR analysis of two independent control knockdown (C) and TAF5-knockdown cells ($n = 3$). **b**, Genome-wide mRNA profiles of TAF5-knockdown versus control knockdown cells. Diagonals indicate 1.5-fold change ($n = 2$, in duplicate). **c**, ChIP analysis of the genes indicated or a control region using antibodies (Ab) to TAF1, TBP or haemagglutinin (HA) as a control (C; $n = 3$). **d**, Activity of the putative *Taf4* enhancer was tested in MEFs, control knockdown (C) or TAF5-knockdown (TAF5 KD) ES cells ($n = 3$). **e**, Quantification of OSKM-induced iPSC formation from OG2 MEFs under TAF-knockdown conditions based on ES-like morphology ($n = 3$). **f**, As **e**, but based on GFP⁺ colonies. * $P < 0.05$, ** $P < 0.01$. Error bars show s.d.

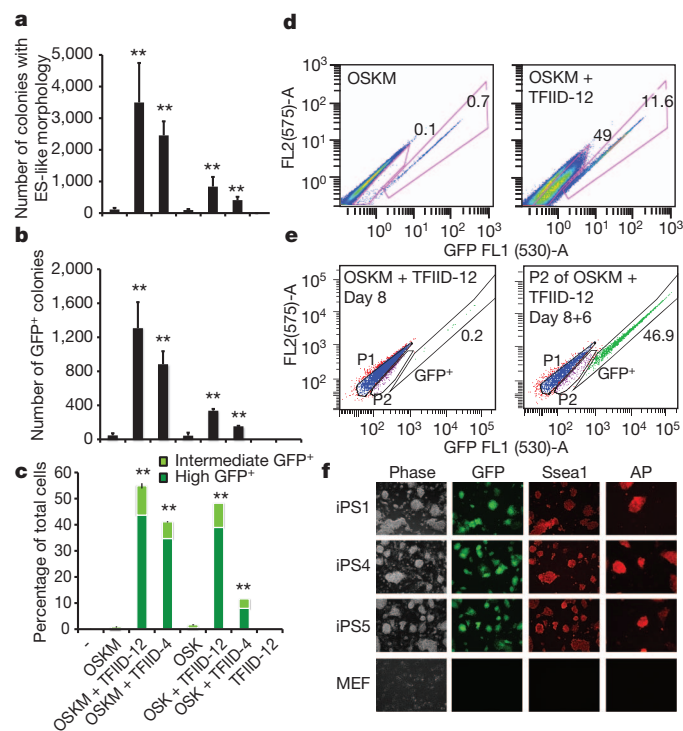


Figure 3 | TFIID potentiates reprogramming to pluripotent cells.

a, Quantification of OSKM- and OSK-induced iPSC formation from OG2 MEFs with TFIID overexpression based on colony morphology ($n = 3$). **b**, As **a**, but based on GFP⁺ colonies. **c**, Quantification of GFP⁺ cells by FACS analysis ($n = 3$). **d**, Representative FACS profiles with intermediate and high GFP⁺ cells enclosed by a box. Fluorescence was recorded at 530 nm (plotted on x-axis) and at 575 nm (y-axis) to distinguish GFP from auto-fluorescence. **e**, The intermediate GFP⁺ population (P2) from OSKM plus TFIID-12 reprogramming was obtained at 8 days after infection by FACS sorting and analysed again after 6 days of culture. The percentage of high GFP⁺ cells is indicated. **f**, OSKM plus TFIID-12 derived iPSCs are GFP, Ssea1 and alkaline phosphatase (AP) positive. iPS1, 4 and 5 indicate clone number. Original magnification, $\times 200$. ** $P < 0.01$. Error bars show s.d.

ChIP sequencing (ChIP-seq) data³ from ES cells indicated binding of Nanog and Oct4 to a region ~10-kb upstream of the mouse *Taf4* gene. This region contains a near-consensus Sox2/Oct4 site, and Nanog and Oct4 binding was confirmed by ChIP-qPCR (Supplementary Fig. 17). The upstream region was tested for enhancer activity in a luciferase-based assay. Figure 2d shows that this region acts as an ES-cell-specific enhancer, which is dependent on TAF5 expression. This indicates that TAF4 is included in the feed-forward regulatory circuitry in ES cells.

Our experiments indicate a requirement for high TFIID levels to maintain the ES cell state. To test whether TFIID also has a role in cellular reprogramming, primary OG2 mouse embryonic fibroblasts (MEFs) were used, which express GFP under the control of the distal *Oct4* enhancer and promoter¹⁶. OSKM-mediated reprogramming under conditions of lentiviral-mediated TAF knockdown showed strongly reduced numbers of colonies with ES-cell-like morphology or GFP fluorescence (Fig. 2e, f). Importantly, transient knockdown (or overexpression, see later) of the tested TAFs did not cause obvious cell death or growth inhibition of MEFs (data not shown). Next, TFIID overexpression was tested using two different pMX-based retroviral mixtures: TFIID-4, containing the core subunits TAF4, 5 and 6 with TBP; or TFIID-12, containing TFIID-4 with TAF3, 4b, 7, 9, 10, 11, 12 and 13. Inclusion of TFIID-12 during OSKM-mediated reprogramming of OG2 MEFs resulted in ~30-fold enhancement in both the number of colonies with ES-cell-like morphology and with GFP fluorescence (Fig. 3a, b). TFIID-4 caused a ~20-fold increase in ES-cell-like/GFP⁺ colonies.

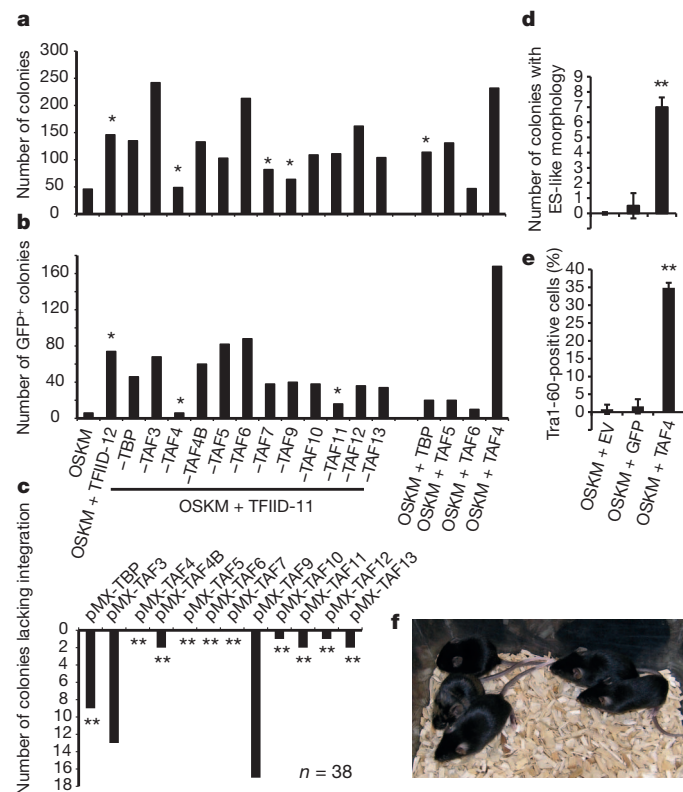


Figure 4 | TAF contributions to reprogramming and developmental potential of iPSCs. **a**, Quantification of iPSC colonies based on morphology. OSKM-induced reprogramming of OG2 MEFs was tested with TFIID-12, TFIID-11 (in which the indicated TAF was omitted), or single TAFs ($n = 2$). **b**, As **a**, but based on GFP⁺ fluorescence. **c**, Genotyping of 38 iPSC clones isolated after OSKM plus TFIID-12 reprogramming. Shown is the number of colonies lacking the indicated subunit. **d**, Reprogramming of human fibroblasts is strongly enhanced by co-expression of TAF4 but not empty vector (EV) or GFP. The number of colonies with ES-like morphology is indicated ($n = 3$). **e**, As **d**, but FACS analysis of Tra1-60⁺ cells ($n = 3$). **f**, Healthy male mice derived from tetraploid complementation with OG2-derived iPSC clone 1. * $P < 0.05$, ** $P < 0.01$. Error bars show s.d.

Omission of *c-Myc* resulted in eight- and fourfold increases caused by TFIID-12 and TFIID-4, respectively. Expression of TFIID subunits in the absence of OSKM failed to induce any ES-cell-like colony. FACS analysis of GFP⁺ cells showed that TFIID-12 and TFIID-4 strongly enhanced the number of both intermediate and high GFP⁺ cells (Fig. 3c, d and Supplementary Fig. 18). Intermediate GFP⁺ cells (P2 population; Fig. 3e) of OSKM plus TFIID-12 reprogramming were isolated by FACS at 8 days after infection, plated for an additional 6 days and analysed again for GFP expression (Fig. 3e). This showed that 47% of the GFP-intermediate cells were converted into high GFP⁺ cells, indicating that the intermediate population may represent unstable and partially reprogrammed cells of which many make it to the pluripotent state. Infection of MEFs with TFIID-12 did not affect cell proliferation, as determined by BrdU incorporation (Supplementary Fig. 19).

Three GFP⁺ colonies from the OSKM plus TFIID-12 transduction were isolated for further characterization. All expressed the ES-cell markers *Ssea1* (also known as *FUT4*) and alkaline phosphatase (Fig. 3f), and contained proviral integrations for OSKM with distinct combinations of 8–9 TFIID subunits (Supplementary Fig. 20). Proviral expression of OSKM and TFIID subunits (Supplementary Figs 21 and 22) was silenced in these clones. Bisulphite sequencing showed demethylation of the *Oct4* promoter (Supplementary Fig. 23). Genome-wide mRNA profiles of these iPSCs strongly resembled ES-cell profiles with high expression of OSKM and TFIID subunits (Supplementary Fig. 24 and Supplementary Table 2).

To obtain further insight into the mechanism of TFIID-stimulated reprogramming, individual TAF retroviruses were left out from the TFIID-12 mixture. Omission of TAF4 abolished the stimulatory effect of TFIID-12, as assessed by morphology and GFP fluorescence (Fig. 4a, b). TAF4 alone stimulated OSKM-mediated reprogramming of OG2 MEFs (Fig. 4a, b) or adult fibroblasts (Supplementary Fig. 25), and TBP and TAF5 had moderate effects. Thirty-eight colonies obtained by OSKM plus TFIID-12 reprogramming were genotyped (Fig. 4c and Supplementary Fig. 26), which revealed integration of TAF4, 5, 6 and 7 retroviruses in all iPSC clones. A genome-wide RNA interference screen for determinants of human ES cells identified TAF2, 7 and 12 (ref. 17). Plotting the results of this study for all TFIID subunits indicates that the combined TAFs have an even more pronounced effect on human ES cells than a set of well-known pluripotency genes (Supplementary Fig. 27). Next, we found that TAF4 overexpression during reprogramming of human fibroblasts by limiting amounts of the OSKM virus strongly stimulated iPSC formation, as assessed by colony formation and Tra1-60 staining (Fig. 4d, e and Supplementary Fig. 28). Together, these analyses indicate that TAF4 has a central role in iPSC formation from mouse as well as human fibroblasts.

Pluripotency of OSKM plus TFIID-12 mouse iPSCs was demonstrated by teratoma formation (Supplementary Fig. 29). Next, OSKM plus TFIID-12 iPSCs were tested for germline transmission by tetraploid complementation, in which diploid iPSCs are the (almost) exclusive source for the embryo. This resulted in normal development of five male mice from iPSC clone 1. Four mice were fully derived from iPSCs as judged from the black coat colour, and one showed a low level of chimaerism (Fig. 4f). Genotyping confirmed the iPSC-clone-1 origin of these mice (Supplementary Fig. 30). After breeding with wild-type females, 12 fetuses were analysed at embryonic day 10.5 and all showed GFP⁺ gonads (Supplementary Fig. 31). Together, these experiments demonstrate pluripotency of TFIID-generated iPSCs.

Our findings reveal that high levels of TFIID are essential to induce and maintain the transcriptional program of pluripotent cells. It is unclear why pluripotent cells have this TFIID dependency. Possibly, the core promoter sequences of key pluripotency genes are inefficient binders of TFIID. Indeed, the *Oct4*, *Sox2* and *c-Myc* promoters lack any TATA sequences and only *Klf4* contains a TATA box (Supplementary Fig. 32). Reduced expression of TFIID by TAF5 or TAF3 (ref. 18) knockdown affected expression of pluripotency genes, but still supported transcription of differentiation genes, which are enriched for

canonical TATA sequences¹⁹ (Fig. 2 and Supplementary Fig. 33). This agrees with studies of *Taf7*, *Taf8* or *Taf10* gene disruptions, which indicate that these TAFs are particularly important in pluripotent cells before gastrulation and less important in differentiated and/or adult tissues^{10,20,21}.

OSKM-induced reprogramming is inefficient^{5,6}, and retroviral co-expression of TFIID TAFs greatly stimulated reprogramming (Fig. 3). Proviral integrations were detected for only eight or nine TAFs, revealing that not all TAFs were required during reprogramming. Partial and alternative TFIID complexes have been detected^{9,10,22} and may support certain transcriptional functions. The TAF4 virus is a common integrant and essential for TFIID-12-stimulated reprogramming (Fig. 4). TAF4 is a structural subunit in the core of TFIID²³ and an ES-cell-specific enhancer is located upstream of the mouse *Taf4* gene. We speculate that TFIID collaborates with pluripotency factors in an auto-stimulatory loop in which high levels of TFIID drive pluripotency genes, which in their turn enhance expression of critical TAFs like TAF4 (Supplementary Fig. 1). This explains the synergism between OSKM and TFIID in reprogramming.

Genome-wide knockdown screens identified TAFs, Mediator subunits and chromatin regulatory factors as critical for pluripotency^{17,22,24–27}. Overexpression of Prmd14 (ref. 17) or Brg1 and Baf155 (ref. 27) also stimulate iPSC formation by the gene-specific OSKM factors. TFIID may share a collective mechanism with these general regulators. We propose that these regulators enable fibroblasts to cross the epigenetic energy barriers for resetting gene expression of the pluripotent state^{5,6}. This predicts that the energetic costs of *trans*-differentiation induced by other gene-specific factors^{2,28} could also be met by overexpression of general regulators like TFIID.

METHODS SUMMARY

Short hairpin RNA (shRNA)-expressing puromycin-resistant lentiviruses were used for knockdown. Quantitative SILAC-based proteomics was performed on TAF6-GFP-expressing ES cells. Genome-wide mRNA expression analysis was performed using Agilent microarrays. ChIP-qPCR was used to analyse promoter occupancies of TFIID. Reprogramming assays were performed using OGG2 MEFs and pMX-based retroviruses. Retroviral integration and silencing were analysed using qPCR. Human fibroblasts were reprogrammed using lentiviruses. Promoter methylation was analysed using bisulphite sequencing. Teratoma formation, blastocyst injections and gonad analysis were performed as described²⁷.

Received 7 November 2011; accepted 30 January 2013.

Published online 17 March 2013.

1. Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437–447 (2008).
2. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
3. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
4. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
5. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49–52 (2009).
6. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).
7. Roeder, R. G. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett.* **579**, 909–915 (2005).
8. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature* **461**, 186–192 (2009).
9. Goodrich, J. A. & Tjian, R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nature Rev. Genet.* **11**, 549–558 (2010).
10. Müller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* **20**, 533–540 (2010).

11. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
12. Juven-Gershon, T. & Kadonaga, J. T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* **339**, 225–229 (2010).
13. Cler, E., Papai, G., Schultz, P. & Davidson, I. Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell. Mol. Life Sci.* **66**, 2123–2134 (2009).
14. Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967–980 (2010).
15. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
16. Szabó, P. E., Hubner, K., Scholer, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* **115**, 157–160 (2002).
17. Chia, N. Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
18. Liu, Z., Scannell, D. R., Eisen, M. B. & Tjian, R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* **146**, 720–731 (2011).
19. Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Rev. Genet.* **8**, 424–436 (2007).
20. Gégonne, A. *et al.* The general transcription factor TAF7 is essential for embryonic development but not essential for the survival or differentiation of mature T cells. *Mol. Cell. Biol.* **32**, 1984–1997 (2012).
21. Voss, A. K. *et al.* Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* **127**, 5449–5461 (2000).
22. Maston, G. A. *et al.* Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *eLife* **1**, e00068 (2012).
23. Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* (2013).
24. Loh, Y. H., Zhang, W., Chen, X., George, J. & Ng, H. H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21**, 2545–2557 (2007).
25. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
26. Lessard, J. A. & Crabtree, G. R. Chromatin regulatory mechanisms in pluripotency. *Annu. Rev. Cell Dev. Biol.* **26**, 503–532 (2010).
27. Singhal, N. *et al.* Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* **141**, 943–955 (2010).
28. Sterneckert, J., Hoing, S. & Scholer, H. R. Oct4 and more: the reprogramming expressway. *Stem Cells* **30**, 15–21 (2011).
29. Munoz, J. *et al.* The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol. Syst. Biol.* **7**, 550 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is funded by the Netherlands Proteomics Centre, Netherlands Organization for Scientific Research (TOP#700.57.302), Deutsche Forschungsgemeinschaft (DFG Priority Program SPP 1356/2) and the European Union (EUTRACC LSHG-CT-2006-037445). We thank M. Groot Koerkamp and D. van Leenen for microarray analysis, M. de Bruijn and O. Kranenburg for advice on lentiviral knockdown experiments, F. Stewart for help with BAC recombineering, I. Davidson and R. Tjian for TAF cDNA constructs, N. Outchkourov for the GFP-tagging vector, B. Roeder, T. Oelgeschläger and G. Kops for antibodies, M. Stehling for help with the FACS analyses, A. Schambach and C. Baum for the lentiviral OSKM vector, M. Sgoddard and T. Cantz for human fibroblasts, M. Radstaak for assistance with human iPSC culture and P. de Graaf, N. Jelluma and M. Vermeulen for critical reading of the manuscript.

Author Contributions W.W.M.P.P. and H.T.M.T. initiated the project. W.W.M.P.P., D.E., H.Z., A.J.R.H., F.C.P.H., H.R.S. and H.T.M.T. designed the experiments. M.P.A.B. and W.W.M.P.P. performed lentiviral knockdown and mRNA analysis. N.M., A.F.M.A., W.W.M.P.P. and A.J.R.H. performed and analysed the mass spectrometry experiments. W.W.M.P.P. and F.C.P.H. performed and supervised the mRNA profiling experiments. A.J.B. and W.W.M.P.P. performed the ChIP and enhancer analysis. W.W.M.P.P., D.E., E.v.d.W., D.W.H., H.v.B., S.M., H.Z. and H.T.M.T. performed the reprogramming experiments. P.L. and K.S. contributed to the statistical analysis. G.W. performed the animal experiments. W.W.M.P.P. and H.T.M.T. wrote the manuscript. All authors approved the manuscript.

Author Information DNA microarray data have been deposited in the public ArrayExpress database under accessions E-TABM-1217 and E-TABM-1218. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.T.M.T. (h.t.m.timmers@umcutrecht.nl).

A syringe-like injection mechanism in *Photorhabdus luminescens* toxins

Christos Gatsogiannis^{1*}, Alexander E. Lang^{2*}, Dominic Meusch¹, Vanda Pfaumann², Oliver Hofnagel¹, Roland Benz³, Klaus Aktories² & Stefan Raunser¹

Photorhabdus luminescens is an insect pathogenic bacterium that is symbiotic with entomopathogenic nematodes¹. On invasion of insect larvae, *P. luminescens* is released from the nematodes and kills the insect through the action of a variety of virulence factors including large tripartite ABC-type toxin complexes² (Tcs). Tcs are typically composed of TcA, TcB and TcC proteins and are biologically active only when complete^{3–5}. Functioning as ADP-ribosyltransferases, TcC proteins were identified as the actual functional components that induce actin-clustering, defects in phagocytosis and cell death^{5–7}. However, little is known about the translocation of TcC into the cell by the TcA and TcB components. Here we show that TcA in *P. luminescens* (TcdA1) forms a transmembrane pore and report its structure in the prepore and pore state determined by cryoelectron microscopy. We find that the TcdA1 prepore assembles as a pentamer forming an α -helical, vuvuzela-shaped channel less than 1.5 nanometres in diameter surrounded by a large outer shell. Membrane insertion is triggered not only at low pH as expected, but also at high pH, explaining Tc action directly through the midgut of insects⁸. Comparisons with structures of the TcdA1 pore inserted into a membrane and in complex with TcdB2 and TccC3 reveal large conformational changes during membrane insertion, suggesting a novel syringe-like mechanism of protein translocation. Our results demonstrate how ABC-type toxin complexes bridge a membrane to insert their lethal components into the cytoplasm of the host cell. We believe that the proposed mechanism is characteristic of the whole ABC-type toxin family. This explanation of toxin translocation is a step towards understanding the host–pathogen interaction and the complex life cycle of *P. luminescens* and other pathogens, including human pathogenic bacteria, and serves as a strong foundation for the development of biopesticides.

To understand ABC-type Tc proteins and, consequently, their mechanism of pore formation and toxin translocation, we used cryoelectron microscopy and single-particle analysis (SPA) to determine the structure of the TcdA1 complex at a resolution of 6.3 Å (Fig. 1 and Supplementary Figs 1 and 2). TcdA1 forms a 17-nm-wide, 25.5-nm-long pentamer with a molecular mass of ~1.41 megadaltons (MDa) (Fig. 1a). The overall bell-like shape of the complex results from an inner, 25.5-nm-long pore that is, for the most part, surrounded by a large outer shell (Fig. 1b). Because the crystal structures of neither full TcA complexes nor subunits are known, we identified all clearly resolved rod-like electron densities and fitted poly-alanine helices into them (Fig. 1c, d). Contrary to previous suggestions⁹ that favoured a β -barrel structure as predicted for the pore-forming anthrax protective antigen¹⁰, the channel shows a clear α -helical architecture with 25 central α -helices. In addition, our computational analysis identified five β -sheets in the upper part of the pore (Fig. 1c). Shaped like a vuvuzela (a flared funnel) that narrows from its maximum diameter of 3 nm to less than 1.5 nm, the central pore is composed of an upper, cone-shaped vestibule

consisting of at least five β -sheets and five α -helices, and a long, tube-shaped channel formed by two consecutive sets of ten α -helices.

Two large domains per subunit of the pentameric complex form the outer shell. Whereas one contains 26 clearly resolved α -helices, the other is dominated by β -sheets (Fig. 1a, d). To locate these regions in the protein sequence, we performed a sequence-based secondary structure prediction analysis of TcdA1 (Supplementary Fig. 3). We identified a large amino-terminal region spanning amino acids 1–1090, composed mainly of α -helices, followed by a large β -strand-dominated region (amino acids 1091–1720), which corresponds well to our structure. In accordance with our structure for the central pore, the carboxy-terminal region of the protein is composed of exceptionally long α -helices (1721–2320) and short terminal β -strands (2324–2516). Thus, each subunit of the TcdA1 pentamer is composed of an N-terminal α -helical domain, a central β -sheet domain and a C-terminal pore-forming domain (Fig. 1d and Supplementary Video 1).

The exceptional length of the helices in the pore-forming domain (up to 90 amino acids) allowed us to assign their corresponding sequence to them precisely (Fig. 1e). We identified a series of four helices per protomer (2020–2284) arranged as pairs of consecutive long and short helices that run antiparallel to each other to form the wall of the tube.

In our structure, the central pore is closed not only by the loops connecting the pore-forming helices, but also by the outer shell that occludes the lower part of the pore. We therefore conclude that our structure represents the prepore state of TcdA1. To obtain the TcdA1 complex in the pore state, we reconstituted the complex into liposomes (Fig. 2a) and determined its structure at a resolution of 19.9 Å using cryoelectron microscopy and SPA (Fig. 2b–e and Supplementary Fig. 2). We note that the umbrella-like shape of the pore complex differs considerably from that of the bell-like prepore structure. Whereas the latter resembles the structure of Yen-Tc K:9 from *Yersinia entomophaga*⁹, the overall shape of the TcdA1 pore complex is similar to the anthrax and α -haemolysin toxin pores^{11,12}.

To provide a better comparison between pore and prepore, we filtered the prepore structure to 19.9 Å and segmented both maps (Fig. 2f, g). When comparing both structures, very large rearrangements of the major domains between the pore and prepore structure become obvious (Fig. 2f, g and Supplementary Fig. 4a, b). The central pore is shifted by 115 Å towards the lower end of the complex, and its lower tubular part penetrates the membrane. Although the resolution is low, the structure of the pore suggests that it is open at this position. This could be achieved by lateral movement of the transmembrane helices during integration into the membrane (Fig. 2g).

The two domains of the outer shell are rotated and shifted relative to each other in a complex manner that results in a more compact subunit with greater distance to adjacent subunits (Supplementary Fig. 4a, b). Consequently, the outer shell has a larger diameter, is open at both sides and encloses mainly the funnel-like part of the central pore, forming a large basin (Supplementary Fig. 4c–e).

¹Department of Physical Biochemistry, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. ²Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, Albert-Ludwigs-Universität Freiburg, 79104 Freiburg, Germany. ³School of Engineering and Science, Jacobs University Bremen, Campusring 1, 28759 Bremen, Germany.

*These authors contributed equally to this work.

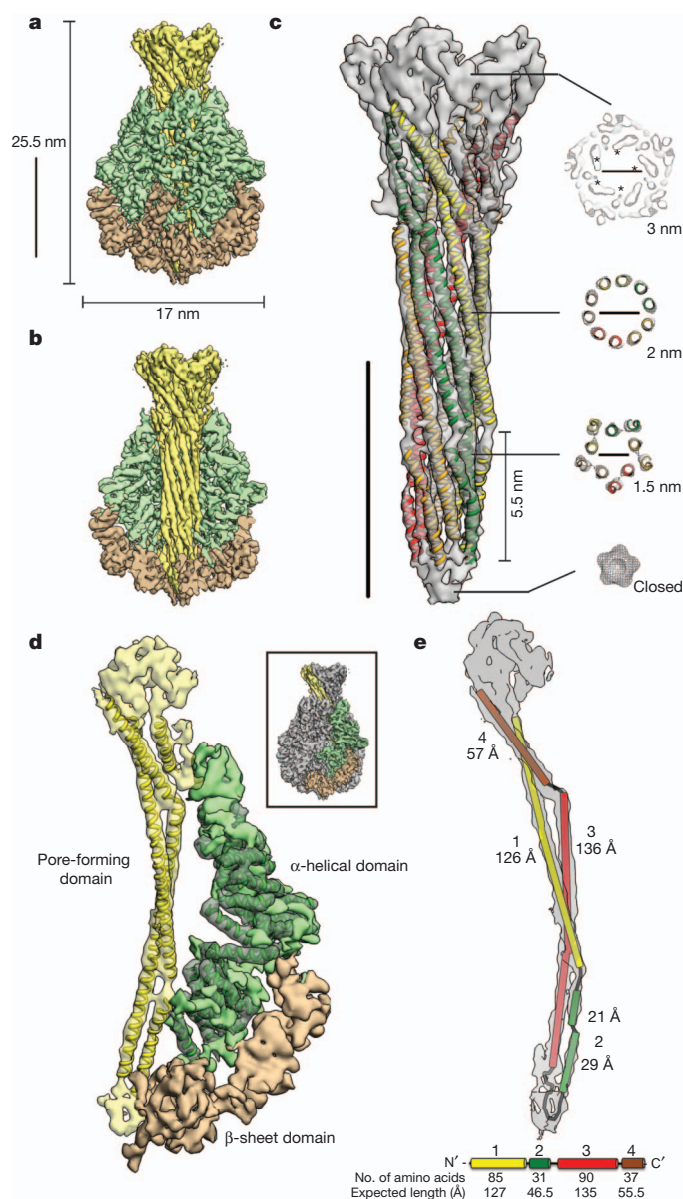


Figure 1 | Cryoelectron microscopy structure of the TcdA1 prepore complex. **a**, Electron density map of the TcdA1 prepore complex. **b**, Cut-away view displaying the central pore. The two domains of the outer shell are respectively depicted in green and brown, and the central pore is depicted in yellow. **c**, Fit of identified α -helices into the electron density of the central pore. At right are shown slices through the central pore at different heights. Asterisks indicate β -sheets. **d**, Extracted electron density of a protomer with identified α -helices. The outlined box shows the localization of the subunit in the complete structure. Densities corresponding to the N-terminal α -helical domain, the central β -sheet domain and the C-terminal pore-forming domain are depicted in green, brown and yellow, respectively. **e**, Sequence assignment to α -helices in the pore-forming domain. The four identified helices are depicted as numbered and coloured cylinders with their respective lengths. At bottom are shown the sequence-based secondary structure prediction for amino acids 2018–2284, the number of amino acids for each predicted helix and the expected lengths of the helices. Helices 1–4 include amino acids 2020–2105, 2109–2140, 2153–2244 and 2249–2285, respectively. Scale bars, 10 nm.

Our data strongly suggest that TcdA1 is an α -helical pore-forming toxin (α -PFT). Whereas other PFTs form the pore during membrane insertion^{13,14}, for TcdA1 the pore is already present before membrane insertion, but is shielded by the outer shell of the protein. Surprisingly, although sequence-based hydrophobicity analyses of TcdA1 predict a higher hydrophobicity for the helices inside the

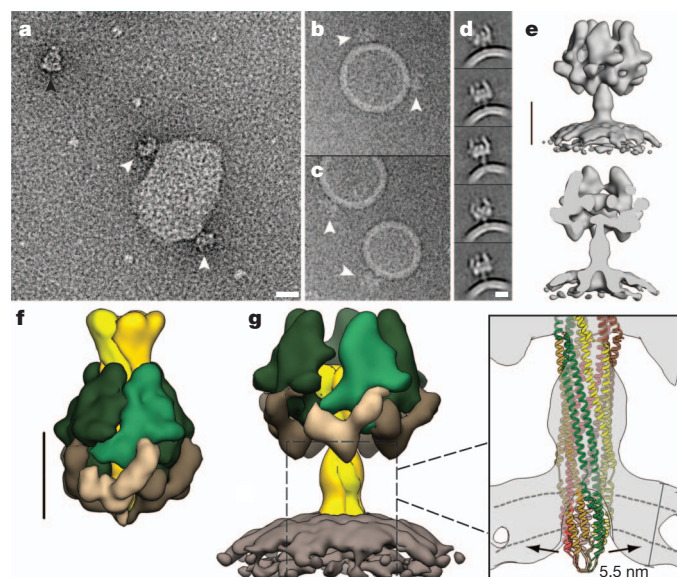


Figure 2 | Cryoelectron microscopy structure of the TcdA1 pore and comparison with the prepore complex. **a–c**, Representative fields of negatively stained (**a**) and vitrified (**b**, **c**) TcdA1 complexes reconstituted in liposomes. Black and white arrowheads indicate molecules in the prepore and pore state, respectively. Scale bars, 20 nm. **d**, Representative class averages of the vitrified reconstituted complex. Scale bar, 10 nm. **e**, Cryoelectron microscopy structure of the TcdA1 pore depicted in side and cut-away views. Scale bar, 10 nm. **f**, **g**, Segmentation of maps of the prepore (**f**) and pore (**g**) states; the two domains of the outer shell are respectively depicted in green and brown, and the central channel is depicted in yellow. The cut-away view in the inset in **g** shows that the transmembrane helices need to move laterally to fit into the density and to open the pore. Scale bar, 10 nm.

membrane (Supplementary Fig. 5a), they do not detect clear transmembrane regions as is the case for other structurally characterized α -PFTs¹⁵. Structure-based calculations of the surface hydrophobicity potential, however, identify strongly hydrophobic stretches in the predicted transmembrane region (Supplementary Fig. 5b), indicating that the hydrophobicity is increased by the ordered assembly of *per se* moderately hydrophobic α -helices.

In analogy to AB-type toxins, such as anthrax toxin¹⁰, the Tc complex is thought to bind first to a receptor on the cell surface before being undergoing endocytosis^{5,9}. On acidification of the endosome, the pH shift induces the insertion of TcA into the endosomal membrane, which translocates either the C component or both the B component and the C component into the cytosol^{5,9}. However, *in vitro*, that is, without receptor and membranes, the prepore TcdA1 complex proved to be stable over a wide pH range (pH 4.5 to pH 10) (Supplementary Fig. 6) and insertion into preformed vesicles was difficult, although not impossible. For more extreme pH values, such as pH 4, we observed many aggregates, but many of the non-aggregated complexes resembled the pore state of the complex (Supplementary Fig. 6a). Surprisingly, a shift to high pH values, such as pH 11, also induced pore formation of TcdA1 *in vitro* (Supplementary Fig. 6e). Depending on protein concentration, we also observed monomers, dimers and tetramers of pentamers (Supplementary Fig. 6e). This effect is probably caused by the interaction of the exposed hydrophobic transmembrane regions and was also observed in other pore-forming toxins¹⁶.

Comparing the structure of this complex obtained by negative-stain electron microscopy and SPA with the negative-stain electron microscopy structure of the prepore complex (Fig. 3a, b) and with the reconstituted complex in liposomes (Fig. 2e) proves that the complex indeed represents the pore state of TcdA1. The observed extreme pH values needed for the transition probably result from the absence of the receptor. Nevertheless, our experiments show that a pH shift induces the transition from the prepore state of TcdA1 to the pore state

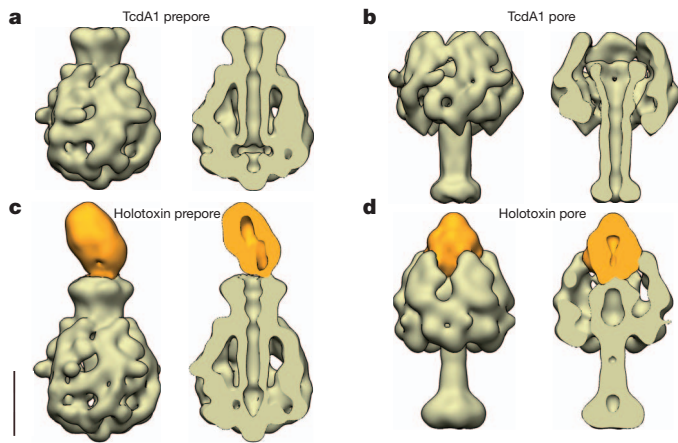


Figure 3 | Structure of the PTC3 holotoxin complex. Side views and central cuts of negatively stained TcdA1 and the PTC3 holotoxin complex in the pre-pore and pore states. **a**, Pre-pore TcdA1; **b**, pore TcdA1; **c**, pre-pore PTC3 holotoxin complex; **d**, pore PTC3 holotoxin complex. TcdA1 and TcdB2–TccC3 are shown in khaki and orange, respectively. Scale bar, 10 nm.

(Supplementary Fig. 7a–d and Supplementary Video 2). As described for many G-protein-coupled receptors, lowering the pH, with resulting protonation of residues, very probably destroys ionic as well as electrostatic interactions and thereby causes conformational changes¹⁷. Our finding that a shift to higher pH values results in the same transition indicates that deprotonation of residues, and not only their protonation, can release the ionic or electrostatic lock that constrains the TcdA1 complex in its pre-pore conformation.

In accordance with our observations for TcdA1 reconstitution in liposomes, lipid bilayer experiments demonstrated that TcdA1 formed ion-permeable channels over a large pH range (pH 4 to pH 11), exhibiting a well-defined single-channel conductance (Supplementary Fig. 8a–c). As expected from our electron microscopy studies, the pore-forming activity of TcdA1 was higher at extreme pH values. At pH 11, the pore-forming activity of TcdA1 was more than 100-fold greater than at pH 6. This probably explains why Tc toxins also directly intoxicate the midgut⁸, which is highly alkaline in many insects¹⁸. However, in nature the nematodes release *P. luminescens* only in the haemocoel¹⁹, where it first colonizes midgut cells, which presumably endocytose the lethal bacterial toxins²⁰. Although it is tempting to speculate that the ability to form a pore at both low and high pH values would allow Tc toxins to intoxicate midgut cells in two ways, from the haemocoel and at the same time from inside the gut, the process in nematodes suggests that this mechanism might not be used in nature.

We measured the TcdA1 pore when reconstituted into a bilayer and, in line with our cryoelectron microscopy structure, found that the TcdA1 pore is funnel shaped with an entrance diameter of 3.0 nm and a minimum diameter of about 1.6 nm (Supplementary Figs 8d and 9 and Supplementary Table 1). This result corroborates our structural data, showing that the central pore of TcdA1 is already formed in the pre-pore state and does not change much during membrane insertion.

To determine where the B and C subunits bind to the A subunit, we first assembled the B and C subunits and then assembled the full PTC3 holotoxin, composed of TcdA1, TcdB2 and TccC3, and determined their structures by electron microscopy and SPA (Fig. 3c, d, Supplementary Fig. 7e–h and 10). The TcdB2 and TccC3 subunits form a strong, dimeric, oval-shaped complex (Supplementary Fig. 10). Corresponding well with biochemical data of the Tc complex from *Xenorhabdus nematophilus*⁴, only one B–C complex rather than five binds to a TcdA1 pentamer directly interacting with the funnel-shaped vestibule of the TcdA1 channel (Fig. 3c). pH-induced pore formation of TcdA1 as described above results in the insertion of the B–C complex into the basin formed by the outer shell and the funnel of the inner channel

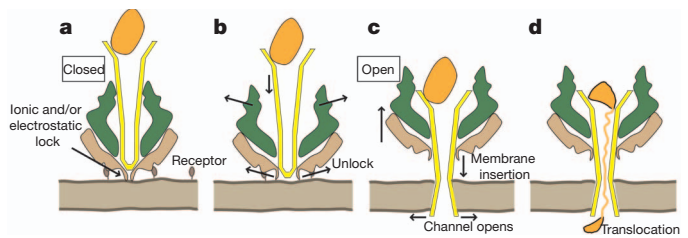


Figure 4 | Syringe-like mechanism for membrane insertion. Model for membrane insertion of the PTC3 complex and translocation of the B–C complex. The two domains of the outer shell are respectively depicted in green and brown, and the central pore of TcdA1 is depicted in yellow. The dimeric complex of TcdB2 and TccC3 is depicted in orange. The beige bar indicates the membrane bilayer. **a**, The PTC3 complex is either taken up orally or secreted into the haemocoel, where it binds to a yet unknown receptor and is endocytosed. **b**, **c**, Lowering the pH in the endosome or increasing the pH in the insect gut destroys an ionic and/or electrostatic lock in the outer shell of the TcdA1 pentamer. The outer shell is then structurally reorganized. As a result, it widens and releases the central pore (**b**), which penetrates the membrane like a syringe (**c**). The pore opens and at the same time the B–C complex is pulled into the basin formed by the funnel of the central pore and the outer shell of the TcdA1 complex. **d**, There, it is unfolded and transported through the central pore.

(Fig. 3d). However, we did not observe its partial or full translocation through the TcdA1 pore, which probably requires a proton or other electrochemical gradient across the membrane, and/or chaperones.

The information provided by these structures allows us to propose the following mechanism for membrane insertion of TcdA1 and protein translocation. First the fully assembled holotoxin composed of TcdA1, TcdB2 and TccC3 is either taken up orally or secreted into the haemocoel of the insect, where it binds to a yet unknown receptor at the cell surface and is endocytosed (Fig. 4a). At that time, the pore is closed and shielded by the outer shell of the protein. A pH-sensitive ionic and/or electrostatic lock opens on acidification of the endosome or on alkalization of the insect gut (Fig. 4b). As a result, the outer shell widens and releases the central pore, which penetrates the membrane like a syringe (Fig. 4c). The pore opens and at the same time the B–C complex is pulled into the basin formed by the central pore and the outer shell, inducing its unfolding and translocation (Fig. 4d and Supplementary Video 3). The structural data on TcdA1 and PTC3 described here reveal that ABC-type toxins are exceptional α -PFTs that insert into membranes using a unique syringe-like mechanism. Hence, it serves as a strong foundation for understanding the action of other pathogens, including human pathogenic bacteria.

METHODS SUMMARY

TcdA1 and TcdB2–TccC3 were recombinantly expressed in *Escherichia coli* BL21-CodonPlus and purified using Ni-IDA affinity as well as size-exclusion chromatography. Membrane insertion and pore activity was measured using black lipid membranes composed of diphytanoyl phosphatidylcholine. For electron microscopy, the samples were applied to carbon-coated or holey-carbon grids either stained with 0.07% uranyl formate or vitrified. Negatively stained complexes were imaged with a JEOL JEM-1400 electron microscope equipped with a LaB₆ cathode operated at 120 kV. Digital electron micrographs were recorded with a 4k × 4k CMOS camera F416 (TVIPS) using minimal-dose conditions. Images of vitrified complexes were made with a JEOL JEM 3200FSC electron microscope equipped with a field-emission gun at an operating voltage of 200 kV. An in-column omega energy filter was used to improve image contrast by zero-loss filtering with a slit width of 15 eV. Micrographs were recorded with an 8k × 8k TemCam-F816 CMOS camera (TVIPS) under minimal-dose conditions (15–20 electrons per Å²). The SPARX software²¹ was used for all image-processing steps with the exception of the initial defocus determination, which was done using CTFIND3²² (Supplementary Figs 2, 11–14). To identify α -helices, we used the programs PHENIX²³ and VOLTRAC²⁴. For visualization and analysis, we used CHIMERA and implemented programs²⁵. Sequence-based secondary structure analysis was carried out using the SABLE protein structure prediction server (<http://sable.cchmc.org/>). A detailed description of our methods can be found in Supplementary Information.

Received 22 September 2012; accepted 5 February 2013.

Published online 20 March 2013.

- Joyce, S. A., Watson, R. J. & Clarke, D. J. The regulation of pathogenicity and mutualism in *Photobacterium*. *Curr. Opin. Microbiol.* **9**, 127–132 (2006).
- ffrench-Constant, R. H. & Bowen, D. J. Novel insecticidal toxins from nematode-symbiotic bacteria. *Cell. Mol. Life Sci.* **57**, 828–833 (2000).
- Waterfield, N. R., Bowen, D. J., Fetherston, J. D., Perry, R. D., & ffrench-Constant, R. H. The *tc* genes of *Photobacterium*: a growing family. *Trends Microbiol.* **9**, 185–191 (2001).
- Sheets, J. J. *et al.* Insecticidal toxin complex proteins from *Xenorhabdus nematophilus*: structure and pore formation. *J. Biol. Chem.* **286**, 22742–22749 (2011).
- Lang, A. E. *et al.* *Photobacterium luminescens* toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science* **327**, 1139–1142 (2010).
- Lang, A. E., Schmidt, G., Sheets, J. J. & Aktories, K. Targeting of the actin cytoskeleton by insecticidal toxins from *Photobacterium luminescens*. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **383**, 227–235 (2011).
- Aktories, K., Lang, A. E., Schwan, C. & Mannherz, H. G. Actin as target for modification by bacterial protein toxins. *FEBS J.* **278**, 4526–4543 (2011).
- Bowen, D. *et al.* Insecticidal toxins from the bacterium *Photobacterium luminescens*. *Science* **280**, 2129–2132 (1998).
- Landsberg, M. J. *et al.* 3D structure of the *Yersinia enterocolitica* toxin complex and implications for insecticidal activity. *Proc. Natl Acad. Sci. USA* **108**, 20544–20549 (2011).
- Young, J. A. T. & Collier, R. J. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* **76**, 243–265 (2007).
- Katayama, H. *et al.* Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles. *Proc. Natl Acad. Sci. USA* **107**, 3453–3457 (2010).
- Song, L. *et al.* Structure of staphylococcal α -hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1865 (1996).
- Geny, B. & Popoff, M. R. Bacterial protein toxins and lipids: pore formation or toxin entry into cells. *Biol. Cell* **98**, 667–678 (2006).
- Parker, M. W. & Feil, S. C. Pore-forming protein toxins: from structure to function. *Prog. Biophys. Mol. Biol.* **88**, 91–142 (2005).
- Iacovache, I., van der Goot, F. G. & Pernot, L. Pore formation: an ancient yet complex form of attack. *Biochim. Biophys. Acta* **1778**, 1611–1623 (2008).
- Eifler, N. *et al.* Cytotoxin ClyA from *Escherichia coli* assembles to a 13-meric pore independent of its redox-state. *EMBO J.* **25**, 2652–2661 (2006).
- Ramon, E. *et al.* Critical role of electrostatic interactions of amino acids at the cytoplasmic region of helices 3 and 6 in rhodopsin conformational properties and activation. *J. Biol. Chem.* **282**, 14272–14282 (2007).
- Dow, J. A. Extremely high pH in biological systems: a model for carbonate transport. *Am. J. Physiol.* **246**, R633–R636 (1984).
- Milstead, J. E. *Heterorhabditis* bacteriophora as a vector for introducing its associated bacterium into the hemocoel of *Galleria mellonella* larvae. *J. Invertebr. Pathol.* **33**, 324–327 (1979).
- ffrench-Constant, R. *et al.* *Photobacterium*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiol. Rev.* **26**, 433–456 (2003).
- Hohn, M. *et al.* SPARX, a new environment for cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
- Rusu, M., Starosolski, Z., Wahle, M., Rigort, A. & Wriggers, W. Automated tracing of filaments in 3D electron tomography reconstructions using Sculptor and Situs. *J. Struct. Biol.* **178**, 121–128 (2012).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to R. S. Goody for continuous support and for comments on the manuscript. We thank I. Vetter for stimulating discussions. This work was supported by the Deutsche Forschungsgemeinschaft grants RA 1781/1-1 (S.R.) and AK 6/22-1 (K.A.) and by Max Planck Society (S.R.).

Author Contributions A.E.L. and V.P. cloned constructs and purified proteins; D.M. performed pH studies; A.E.L. and R.B. performed and analysed black lipid bilayer assays; C.G. screened samples and collected negative-stain electron microscopy data; C.G. and O.H. collected cryoelectron microscopy data; C.G. processed and refined electron microscopy data; C.G. and S.R. analysed electron microscopy data; K.A. and S.R. designed the study. C.G. and S.R. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The coordinates for the electron microscopy structures have been deposited in EMDDataBank under accession codes EMD-2297 to EMD-2301. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (stefan.raunser@mpi-dortmund.mpg.de).

Critical role of *Trib1* in differentiation of tissue-resident M2-like macrophages

Takashi Satoh^{1,2}, Hiroyasu Kidoya³, Hisamichi Naito³, Masahiro Yamamoto^{4,5}, Naoki Takemura^{1,2}, Katsuhiko Nakagawa^{1,2}, Yoshichika Yoshioka⁶, Eiichi Morii⁷, Nobuyuki Takakura³, Osamu Takeuchi^{1,2,8} & Shizuo Akira^{1,2}

Macrophages consist of at least two subgroups, M1 and M2 (refs 1–3). Whereas M1 macrophages are proinflammatory and have a central role in host defence against bacterial and viral infections^{4,5}, M2 macrophages are associated with responses to anti-inflammatory reactions, helminth infection, tissue remodelling, fibrosis and tumour progression⁶. *Trib1* is an adaptor protein involved in protein degradation by interacting with COP1 ubiquitin ligase⁷. Genome-wide association studies in humans have implicated *TRIB1* in lipid metabolism^{8–10}. Here we show that *Trib1* is critical for the differentiation of F4/80⁺MR⁺ tissue-resident macrophages—that share characteristics with M2 macrophages (which we term M2-like macrophages)—and eosinophils but not for the differentiation of M1 myeloid cells. *Trib1* deficiency results in a severe reduction of M2-like macrophages in various organs, including bone marrow, spleen, lung and adipose tissues. Aberrant expression of C/EBP α in *Trib1*-deficient bone marrow cells is responsible for the defects in macrophage differentiation. Unexpectedly, mice lacking *Trib1* in haematopoietic cells show diminished adipose tissue mass accompanied by evidence of increased lipolysis, even when fed a normal diet. Supplementation of M2-like macrophages rescues the pathophysiology, indicating that a lack of these macrophages is the cause of lipolysis. In response to a high-fat diet, mice lacking *Trib1* in haematopoietic cells develop hypertriglyceridaemia and insulin resistance, together with increased proinflammatory cytokine gene induction. Collectively, these results demonstrate that *Trib1* is critical for adipose tissue maintenance and suppression of metabolic disorders by controlling the differentiation of tissue-resident M2-like macrophages.

Members of the tribble family are pseudokinase proteins that are conserved among species and implicated in various human diseases, such as leukaemia and metabolic disorders. Tribble proteins interact with an E3 ubiquitin ligase, COP1, for protein degradation. In thioglycollate-elicited macrophages, *Trib1* is important for interleukin (IL)-12p40 production to lipopolysaccharide¹¹. In addition, *Trib1* and *Trib2* have been implicated in acute myeloid leukaemia^{12,13}. *Trib3* can inhibit insulin signalling¹⁴, although *Trib3*^{-/-} mice do not show defects in insulin signalling or glucose homeostasis¹⁵. However, the roles of tribble proteins in haematopoietic cell development have not been clarified.

First, we examined the populations of tissue-resident macrophages in mice lacking tribble family genes in the spleen. Although the proportions of B cells, T cells, dendritic cells and Ly6C^{high}Mac1⁺ inflammatory monocytes were not altered between wild-type and *Trib1*^{-/-} splenocytes, F4/80⁺Mac1⁺ macrophages, which also expressed Mrc1 (also called MR), *Arg1* and *Fizz1* (also called *Retnla*), were markedly reduced and Siglec-F⁺CCR3⁺ eosinophils were absent in *Trib1*^{-/-} spleens (Fig. 1a and Supplementary Figs 1 and 2). In contrast, the neutrophil population was increased in *Trib1*^{-/-} spleens (Fig. 1a and

Supplementary Fig. 3). Splenic F4/80⁺ red pulp macrophages phagocytose aged red blood cells and accumulate iron^{16,17}. Immunohistochemical staining of spleen sections confirmed the absence of red pulp macrophages in *Trib1*^{-/-} spleen (Fig. 1b). Furthermore, Perl's Prussian blue staining of the spleen sections revealed that iron did not accumulate in *Trib1*^{-/-} mice (Fig. 1c and Supplementary Fig. 4). In contrast, splenic metallophilic and marginal zone macrophages were normal in *Trib1*^{-/-} mice (Supplementary Fig. 5). In addition to the spleen, tissue-resident macrophages in other tissues were severely decreased in *Trib1*^{-/-} mice; however, peritoneal resident macrophages were comparable between wild-type and *Trib1*^{-/-} mice (Supplementary Fig. 6). Newly generated *Trib2*^{-/-} and *Trib3*^{-/-} mice did not show any defects in myeloid and lymphoid cells in the spleen (Supplementary Figs 7 and 8). Because MR, *Arg1* and *Fizz1* expression is a hallmark characteristic of M2 macrophages, we termed this population M2-like macrophages. Thus, these findings indicate that *Trib1* is critical for the differentiation of tissue-resident M2-like macrophages and eosinophils in the peripheral organs.

Differentiation of haematopoietic cells, including macrophages, occurs in the bone marrow, followed by their migration to peripheral tissues via the bloodstream. Consistent with the defects observed in peripheral organs, numbers of F4/80⁺Mac1⁺ cells and Siglec-F⁺CCR3⁺ eosinophils were severely decreased in the blood and bone marrow cells from *Trib1*^{-/-} mice whereas numbers of Gr-1^{high} neutrophils were slightly increased (Fig. 1d, e). However, inflammatory monocytes were comparable between wild-type and *Trib1*^{-/-} bone marrow cells (Supplementary Fig. 9). The adoptive transfer of *Trib1*^{-/-} bone marrow cells to wild-type mice failed to increase F4/80⁺ macrophage numbers in the bone marrow and spleen (data not shown). Furthermore, competitive transfer of CD45.1⁺ wild-type and CD45.2⁺ *Trib1*^{-/-} bone marrow cells (1:1 ratio) to sublethal-irradiated wild-type mice led to severely impaired development of CD45.2⁺ macrophages and eosinophils, and increased the population of neutrophils (Supplementary Fig. 10). These findings demonstrate that the defects observed in *Trib1*^{-/-} mice are intrinsic to haematopoietic cells, and that *Trib1* is critical for regulating the proper differentiation of myeloid cells in the bone marrow. To delineate the developmental competency of *Trib1*^{-/-} bone marrow cells, we performed colony-forming assays. Whereas granulocyte/neutrophil colonies were increased, macrophage colonies were severely decreased and eosinophil colonies were not generated in *Trib1*^{-/-} bone marrow cells compared with wild-type cells (Fig. 2a). We found that macrophage colonies could be morphologically classified into two subgroups, namely aggregated/small and diffused/large macrophages (Fig. 2b). Although most wild-type macrophage colonies were aggregated and small, the macrophage colonies obtained from *Trib1*^{-/-} bone marrow cells were diffused and large (Fig. 2b). These results indicated that *Trib1* is essential for the proper

¹Laboratory of Host Defense, WPI Immunology Frontier Research Center (WPI IFReC), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ²Department of Host Defense, Research Institute for Microbial Diseases (RIMD), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ³Department of Signal Transduction, Research Institute for Microbial Diseases (RIMD), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁴Laboratory of Immunoparasitology, WPI Immunology Frontier Research Center (WPI IFReC), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁵Department of Immunoparasitology, Research Institute for Microbial Diseases (RIMD), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁶Laboratory of Biofunctional Imaging, WPI Immunology Frontier Research Center (WPI IFReC), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁷Department of Pathology, Graduate School of Medicine, Osaka University, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan. ⁸Laboratory of Infection and Prevention, Institute for Virus Research, Kyoto University, 53 Shogoin Kawara-cho, Sakyo-ku, Kyoto 606-8507, Japan.

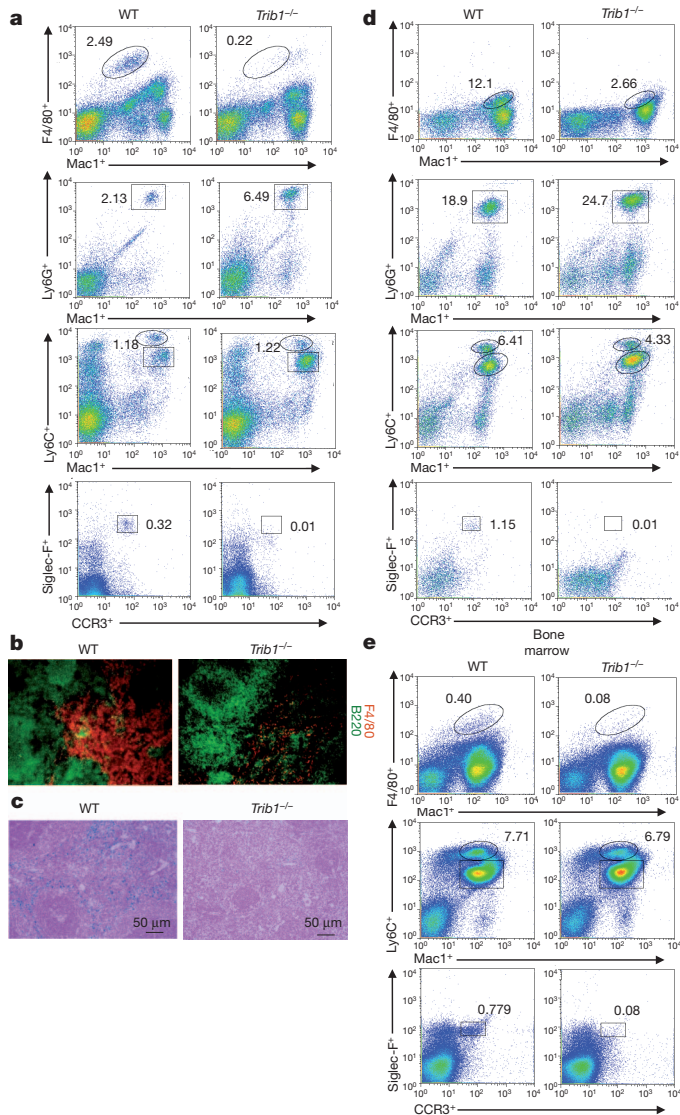


Figure 1 | Lack of tissue-resident M2-like macrophages and eosinophils in *Trib1*-deficient mice. **a**, Flow cytometric analyses of splenocytes. The expression levels of F4/80⁺ and Mac1⁺ tissue-resident M2-like macrophages in splenocytes are shown (top). The proportions of inflammatory monocytes and neutrophils (middle) and eosinophils (bottom) in splenocytes are also shown. Similar results were obtained in eight independent experiments. **b**, Wild-type (WT) and *Trib1*^{-/-} spleen sections were stained for F4/80 (red) and B220 (green). **c**, Images of Perl's Prussian blue staining for ferric iron in the spleens of wild-type and *Trib1*^{-/-} mice. Scale bars, 50 μ m. **d**, Flow cytometric analysis of cell populations in the blood. The percentages of F4/80⁺ and Mac1⁺ cells in the blood are shown. **e**, Flow cytometric analysis of bone marrow. The percentages of F4/80⁺ and Mac1⁺ tissue-resident M2-like macrophages in the bone marrow are shown (top). The proportions of inflammatory monocytes and neutrophils (middle) and eosinophils (bottom) in the bone marrow are also shown. Similar results were obtained in three independent experiments (b–e).

differentiation of bone marrow myeloid cells. Proportions of CMP and CLP as well as GMP were comparable between wild-type and *Trib1*^{-/-} mice (see Supplementary Fig. 11 for definitions). Furthermore, the expression pattern of transcriptional factors, which are involved in myeloid cell differentiation, in GMP and MDP was also comparable between wild-type and *Trib1*^{-/-} mice (Supplementary Fig. 12), suggesting that *Trib1* controls myeloid cell differentiation downstream of GMP.

Retroviral expression of full-length *Trib1* in *Trib1*^{-/-} bone marrow cells resulted in a marked increase in aggregated and small macrophage colonies and eosinophil colonies, but decreased granulocyte/

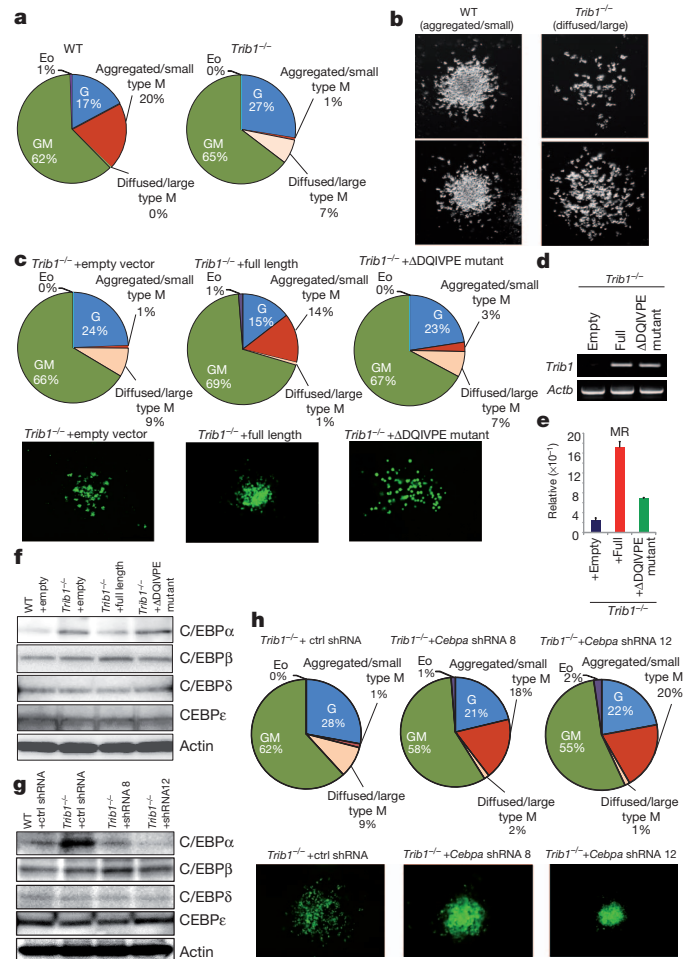


Figure 2 | *Trib1* controls macrophage, eosinophil and neutrophil differentiation via the ubiquitin-binding domain. **a**, **b**, Bone marrow cells were subjected to colony-forming assays. Each colony was counted depending on the morphology (**a**), and images of the cell types of macrophage colonies are shown (**b**). Eo, eosinophil colonies; G, granulocyte/neutrophil colonies; GM, granulocyte/macrophage colonies; M, macrophage colonies. **c–e**, Role of the COP1-binding site in macrophage differentiation. Bone marrow cells lacking *Trib1* were infected with retroviruses expressing empty vector, full-length *Trib1* (mTrib1) or *Trib1*(Δ DQIVPE) mutant. Colony-forming assays were performed in retrovirus-infected (GFP⁺) cells. Images of the retrovirus-infected (GFP⁺) cell types of macrophage colonies are shown in **c**. **d**, Semi-quantitative PCR analyses were performed to measure the expression of *Trib1* in empty vector, full-length *Trib1* (full length) or mutant *Trib1*(Δ DQIVPE) in macrophage colonies. **e**, Quantitative PCR analyses were performed for MR expression in macrophage colonies expressing empty vector, *Trib1* (full length) or mutant *Trib1*(Δ DQIVPE). Error bars indicate s.d. of duplicates. Similar results were obtained in three independent experiments (**a–e**). **f**, Immunoblot analyses of various C/EBP family molecules. Lineage-negative fractions from wild-type, *Trib1*^{-/-} cells, *Trib1*^{-/-} cells expressing full-length and *Trib1*^{-/-} cells expressing mutant *Trib1*(Δ DQIVPE) were analysed by western blotting. **g**, **h**, Role of C/EBP α in macrophage differentiation. Bone marrow cells lacking *Trib1* were infected with retroviruses expressing empty vector, *Cebpa* shRNA 8 or *Cebpa* shRNA 12. Immunoblot analyses of various C/EBP family molecules. Lineage-negative fractions from wild-type, *Trib1*^{-/-} cells and *Trib1*^{-/-} cells expressing *Cebpa* shRNA 8 or *Cebpa* shRNA 12 were analysed by western blotting (**g**). Colony-forming assays were performed in infected (GFP⁺) cells, and images of the retrovirus-infected (GFP⁺) cell types of macrophage colonies are shown (**h**). Similar results were obtained in two independent experiments (**f–h**).

neutrophil colonies (Fig. 2c, d and Supplementary Fig. 13). In contrast, expression of *Trib1* lacking the COP1-binding site (*Trib1*(Δ DQIVPE) mutant) in *Trib1*^{-/-} bone marrow cells failed to restore the differentiation defects. Moreover, expression of full-length *Trib1*, but not of

Trib1(Δ DQIVPE) mutant, restored expression of M2 macrophage marker genes such as *MR* (Fig. 2e). Next, we investigated the expression of potential target proteins of the Trib1–COP1 axis in *Trib1*^{−/−} bone marrow cells. Among the transcription factors, C/EBP family members are important for determining the balance between granulopoiesis and monoopoiesis¹⁸. We found that expression of C/EBP α , but not other transcriptional factors involved in myeloid cell differentiation, was increased in lineage-negative bone marrow cells and

macrophage colonies lacking Trib1 (Fig. 2f and Supplementary Fig. 14). The expression of full-length Trib1 in *Trib1*^{−/−} bone marrow cells suppressed the level of C/EBP α expression (Fig. 2f). To address whether increased expression of C/EBP α is involved in aberrant myeloid cell differentiation under Trib1 deficiency, we genetically inhibited *Cebpa* in *Trib1*^{−/−} bone marrow cells using short hairpin RNAs (shRNAs) (Supplementary Fig. 15). The decreased C/EBP α protein expression in *Trib1*^{−/−} bone marrow cells was similar to levels of

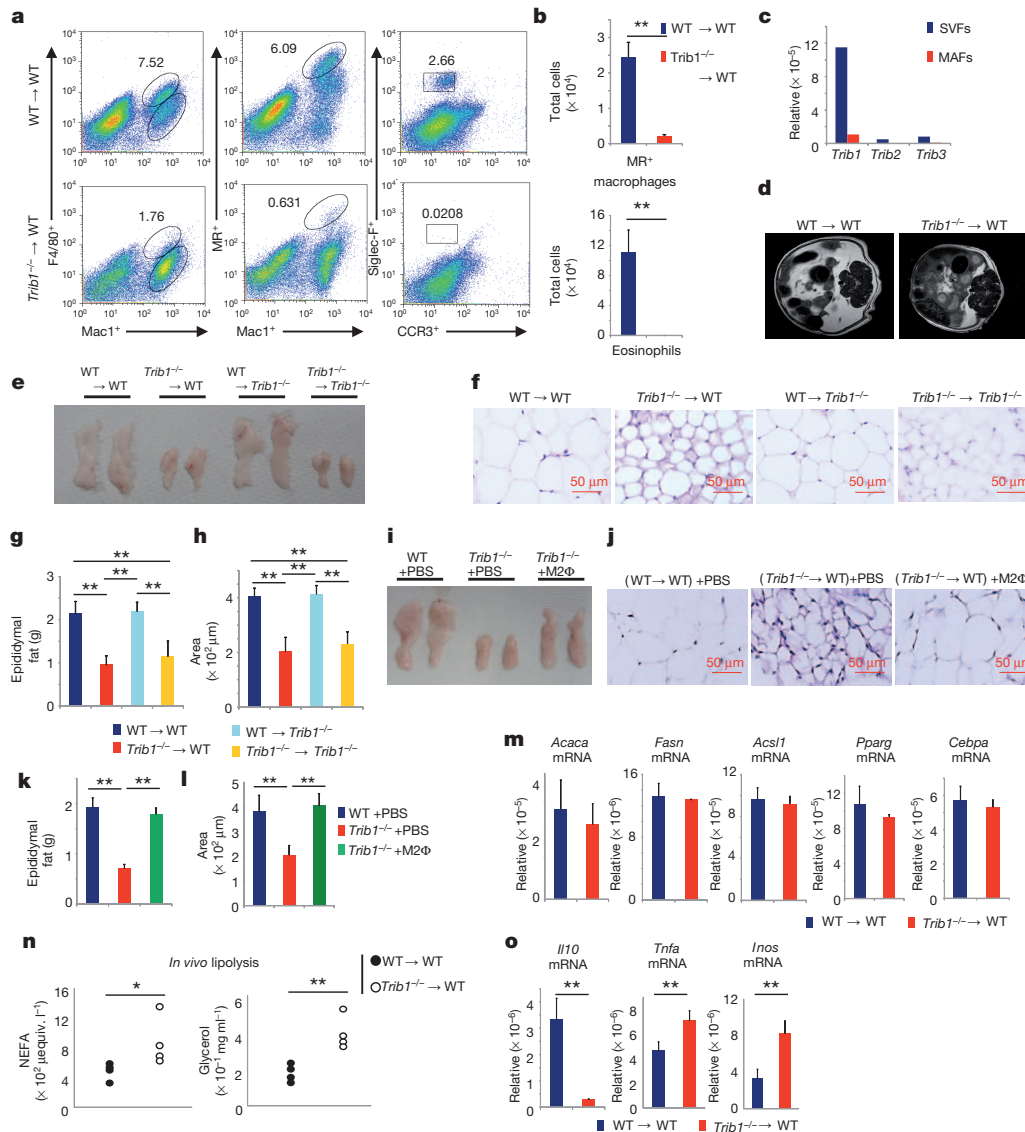


Figure 3 | Impact of Trib1 deficiency on lipodystrophy. **a**, **b**, Flow cytometric analyses of haematopoietic cells in adipose tissue. The expression levels of F4/80⁺ and Mac1⁺ or MR⁺ and Mac1⁺ tissue-resident M2 macrophages in adipose tissue (left and centre) and Siglec-F⁺ and CCR3⁺ eosinophils (right) are shown. The total numbers of MR⁺ macrophages and eosinophils in adipose tissues from wild-type and *Trib1*^{−/−} chimaeric mice fed a normal diet are shown (**b**). Error bars indicate s.d. of duplicates. Similar results were obtained in three independent experiments. **c**, Expression levels of *Trib1*, *Trib2* and *Trib3* in SVFs and MAFs, assessed by quantitative PCR. Similar results were obtained in two independent experiments. **d**, The epididymal adipose tissues from wild-type and *Trib1*^{−/−} chimaeric mice fed a normal diet were analysed by magnetic resonance imaging (MRI). **e**, Images of the epididymal adipose tissues from each indicated chimaeric mice fed a normal diet are shown. **f**, Haematoxylin and eosin-stained images of paraffin-embedded epididymal adipose tissue sections from each indicated chimaeric mice fed a normal diet. Scale bars, 50 μ m. **g**, **h**, The total epididymal fat weight was measured (**g**) and each area of adipocytes was measured (**h**) in each indicated chimaeric mouse. Error bars indicate s.d. of duplicates. **i**, Images of

the epididymal adipose tissues from each indicated chimaeric mice fed a normal diet are shown. **j**, Haematoxylin-and-eosin-stained images of paraffin-embedded epididymal adipose tissue sections from each indicated chimaeric mice fed a normal diet. Scale bars, 50 μ m. **k**, **l**, The total epididymal fat weight (**k**) and each area of adipocytes were measured (**l**) in each indicated chimaeric mouse. Error bars indicate s.d. of duplicates. **m**, Total RNA was prepared from wild-type and *Trib1*^{−/−} chimaeric mice fed a normal diet. The expression levels of mRNAs encoding the indicated proteins were determined by quantitative PCR. Error bars indicate s.d. of duplicates. Similar results were obtained in three independent experiments (**e**–**m**). **n**, Lipolysis assays in wild-type and *Trib1*^{−/−} mice fed a normal diet. The serum glycerol and NEFA levels were measured *in vivo*. Similar results were obtained in two independent experiments. **o**, Total RNA was prepared from wild-type and *Trib1*^{−/−} chimaeric mice fed a normal diet. The expression levels of mRNAs encoding the indicated proteins were determined by quantitative PCR. Error bars indicate s.d. of duplicates. Similar results were obtained in three independent experiments. Statistical significance in **b**, **g**, **h**, **k**, **l**, **n**, **o** was determined using the Student's *t*-test. **P* < 0.05; ***P* < 0.01.

wild-type cells (Fig. 2g). Repression of *Cebpa* resulted in an increase in the aggregated and small macrophage colonies and eosinophil colonies, but a decrease in granulocyte/neutrophil colonies (Fig. 2h). Taken together, these data demonstrate that Trib1 regulates myeloid cell differentiation by altering the expression of C/EBP α in a COP1-dependent manner.

Recent genome-wide association studies revealed that genetic variants in the loci corresponding to *TRIB1* are associated with increased plasma lipoprotein concentrations, a high risk of ischaemic heart disease and myocardial infarction in humans^{8–10}. These studies prompted us to investigate the role of Trib1 in the adipose tissue and liver of mice. Because we did not observe histological changes in the liver in the absence of Trib1 in mice fed normal chow (Supplementary Fig. 16), we next examined the adipose tissues. Consistent with observations in other organs, MR⁺F4/80⁺ adipose-tissue-resident M2-like macrophages were severely decreased in the stromal vascular fraction (SVF) derived from *Trib1*^{−/−} epididymal adipose tissues compared with wild-type mice (Supplementary Fig. 17). Mice lacking Trib1 in haematopoietic cells also showed a reduced number of M2-like macrophages and lacked eosinophils in epididymal adipose tissues (Fig. 3a, b), and *Trib1* mRNA was rarely expressed in the normal mature adipocyte fraction (MAF) (Fig. 3c), indicating that the defect is intrinsic to haematopoietic cells. In contrast, CD11c⁺Mac1⁺ M1 macrophages were hardly present in wild-type and *Trib1*^{−/−} adipose tissues from mice fed a normal chow diet (Supplementary Fig. 17). Unexpectedly, MRI analyses revealed that Trib1 deficiency in haematopoietic cells severely reduced the abdominal adipose tissues (Fig. 3d). The epididymal adipose tissues and size of each epididymal adipocyte were significantly smaller in mice lacking Trib1 in all tissues as well as in littermate wild-type mice fed a normal diet (Fig. 3e–h). Thus, the mice showed a lipodystrophic phenotype. In contrast, *Trib1*^{−/−} mice harbouring wild-type haematopoietic cells did not develop lipodystrophy (Fig. 3e, f). These findings demonstrate that Trib1 deficiency in haematopoietic cells is responsible for development of this pathophysiology. To investigate whether loss of M2-like macrophages is responsible for the development of lipodystrophy, we supplemented *Trib1*^{−/−} mice with these macrophages (Supplementary Fig. 18). After 3 weeks, the epididymal adipose tissues and the size of each adipocyte from *Trib1*^{−/−} mice reconstituted with these macrophages were significantly larger than in PBS-treated *Trib1*^{−/−} mice (Fig. 3i–l). Collectively, these data indicate that lack of M2-like macrophages, caused by Trib1 deficiency, is important for the development of the lipodystrophic phenotype.

Next, we examined the mechanisms of the lipodystrophy caused by Trib1 deficiency in haematopoietic cells. Adipose tissues are maintained by balance between lipogenesis and lipolysis. The mRNA levels of genes involved in adipocyte differentiation and lipogenesis were comparable between wild-type and *Trib1*^{−/−} bone marrow chimaeric mice (Fig. 3m and Supplementary Fig. 19; see Methods). In contrast, the serum levels of non-esterified fatty acids (NEFAs) and glycerol were significantly elevated in *Trib1*^{−/−} bone marrow chimaeric mice (Fig. 3n), indicative of enhanced lipolysis caused by Trib1 deficiency. Cytokines produced by macrophages contribute to adipose tissue inflammation and lipolysis. Expression of IL-10 was severely reduced in epididymal adipose tissues lacking Trib1 in haematopoietic cells. In contrast, *Tnf* and *Inos* mRNA expression were elevated in adipose tissues from *Trib1*^{−/−} bone marrow chimaeric mice (Fig. 3o). Because several reports indicated that IL-10 has a central role in repressing lipolysis in adipocytes in addition to having anti-inflammatory effects^{19–21}, it is possible that M2-like macrophages are important for maintaining adipose tissues, at least in part, through the production of IL-10. However, the contribution of additional molecules produced by the macrophages may also be critical for amelioration of lipodystrophy.

Lipodystrophy is often associated with metabolic abnormalities. Although significant differences in levels of the parameters such as

serum glucose, cholesterol, triglyceride and insulin were not detected between wild-type and *Trib1*^{−/−} bone marrow chimaeric mice fed a normal diet, the feeding of mice with a high-fat diet (HFD) led to much higher elevations of these parameters in the serum of *Trib1*^{−/−} bone marrow chimaeric mice (Fig. 4a). Mice lacking Trib1 throughout the whole body and in haematopoietic cells also developed glucose intolerance and insulin resistance on a HFD (Fig. 4b, c and Supplementary Fig. 20). In contrast, body weight did not differ significantly between wild-type and *Trib1*^{−/−} bone marrow chimaeric mice fed a HFD (Supplementary Fig. 21). Whereas numbers of M1-type macrophages increased in both wild-type and *Trib1*^{−/−} bone marrow chimaeric mice as well as mice lacking Trib1 in the whole body fed a HFD,

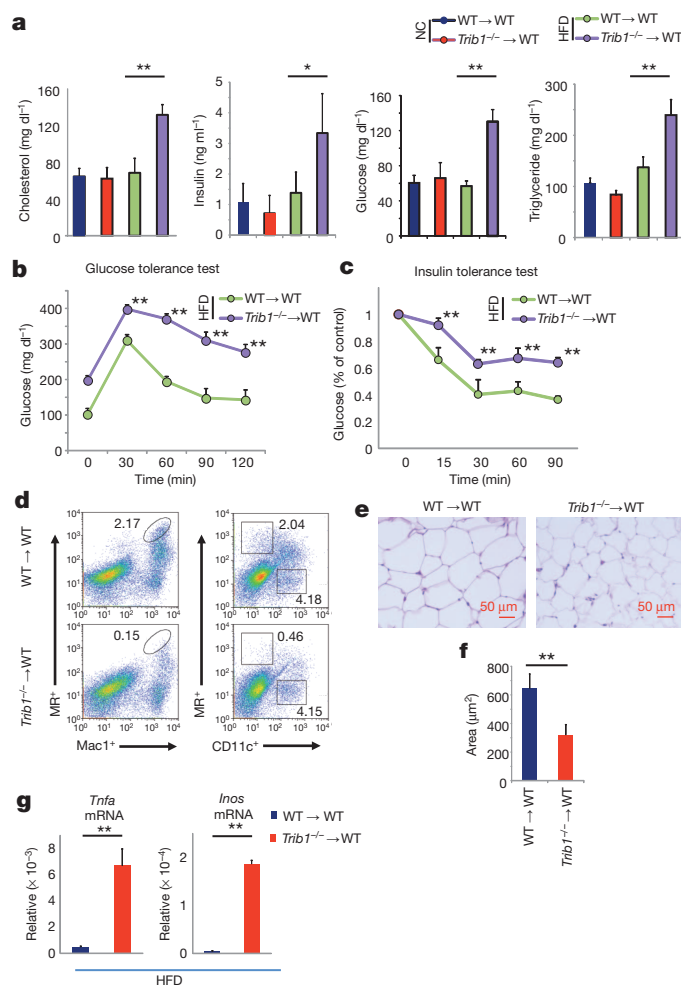


Figure 4 | Mice lacking Trib1 in haematopoietic cells develop exacerbated metabolic disorders on a high-fat diet. **a**, Serum samples were obtained from bone-marrow-transferred wild-type and *Trib1*^{−/−} littermates fed normal chow (NC) or a high-fat diet (HFD). The concentrations of cholesterol, insulin, glucose and triglyceride were measured. Error bars indicate s.d. of duplicates. **b**, **c**, Fasting male bone-marrow-transferred wild-type and *Trib1*^{−/−} littermates fed a HFD were challenged with intraperitoneal glucose (**b**) or insulin (**c**), and blood was sampled for glucose analyses at the indicated times. Data are mean s.d. of 5 samples. **d**, Flow cytometric analyses of haematopoietic cells from adipose tissues in wild-type and *Trib1*^{−/−} mice fed a HFD. The expression levels of MR⁺ and Mac1⁺ tissue-resident macrophages are shown (left). The MR⁺ and CD11c⁺ macrophages in adipose tissue are also shown (right). **e**, **f**, Haematoxylin-and-eosin-stained images of paraffin-embedded epididymal adipose tissue sections are shown (**e**), and each area of adipocytes was measured (**f**). Scale bars, 50 μ m. **g**, Total RNA was prepared from wild-type and *Trib1*^{−/−} chimaeric mice fed a HFD. Error bars indicate s.d. of duplicates. The expression levels of mRNAs encoding the indicated genes were determined by quantitative PCR. Error bars indicate s.d. of duplicates. Statistical significance in **a–c**, **f**, **g** was determined using the Student's *t*-test. **P* < 0.05; ***P* < 0.01.

M2-like macrophages were severely decreased in the absence of *Trib1* (Fig. 4d and Supplementary Fig. 22). *Trib1*^{-/-} chimaeric mice had smaller adipocytes on a HFD (Fig. 4e, f). Notably, *Tnf* and *Inos* mRNA were markedly augmented in *Trib1*^{-/-} chimaeric mice (Fig. 4g). Taken together, these findings suggest that *Trib1*^{-/-} bone marrow chimaeric mice develop metabolic disorders on a HFD through impaired lipid buffering caused by lipodystrophy.

Although we previously showed that *Trib1* regulates C/EBP β expression in thioglycollate-elicited macrophages¹¹, we found that M2-like macrophages were present in *Cebpb*^{-/-} mice (data not shown). Thus, *Trib1* seems to control different target proteins for degradation depending on the cell types and their differentiation stages. We previously showed that *Jmjd3* is essential for M2-like macrophage differentiation in response to chitin administration and helminth infection^{22,23}. However, *Jmjd3* deficiency did not show any defect in tissue-resident M2-like macrophage generation (Supplementary Fig. 23). Reciprocally, *Trib1* deficiency did not alter the activation of M2-like macrophages to chitin administration (Supplementary Fig. 24). It can be assumed that macrophage subtypes are more complex *in vivo*, although the precise classification of M2-type macrophages *in vivo* is yet to be clarified.

Chronic low-grade inflammation in adipose tissues is thought to have a central role in the exacerbation of metabolic disorders^{24,25}. M1-type macrophages infiltrate into obese adipose tissues via MCP1, where they are activated in part by saturated fatty acids and cause low-grade inflammation by producing proinflammatory cytokines^{26,27}. However, tissue-resident M2-like macrophages, which differentiate depending on the presence of *Trib1* in the bone marrow, are subsequently activated in adipose tissues and the liver in response to IL-4 and IL-13 produced from eosinophils by a PPAR- γ - and PPAR- δ -dependent mechanism, respectively^{21,28–30}. Further studies are required to clarify how tissue-resident M2-like macrophages maintain adipose tissue homeostasis. *Trib1* was also essential for the development of eosinophils; nevertheless, we believe that *Trib1* maintains adipose tissues by controlling M2-like macrophages, as supplementation of wild-type macrophages alone could rescue the lipodystrophy observed in *Trib1*^{-/-} mice.

Mutations in *TRIB1* have been implicated in metabolic disorders of humans. It is tempting to speculate that *TRIB1* functions in a similar manner in humans and that other pathophysiologicals, such as tumour progression, angiogenesis or tissue remodelling, are affected by these tissue-resident M2-like macrophages.

METHODS SUMMARY

Mice, reagents, cells and plasmids. Full details are provided in the Methods.

Colony-forming assay. Bone marrow cells (4×10^4) were seeded on 3.5-cm dishes and cultured for 7 days with Methocult (no. 3534; Stem Cell Technologies) containing SCF, IL-6 and IL-3 in the presence of 20 ng ml^{-1} M-CSF. The numbers of GFP-positive cells were counted for colony-forming assays or collected for real-time (q)PCR and western blotting.

Full Methods and any associated references are available in the online version of the paper.

Received 19 March 2012; accepted 22 January 2013.

Published online 20 March 2013.

- Martinez, F. O., Helming, L. & Gordon, S. Alternative activation of macrophages: an immunologic functional perspective. *Annu. Rev. Immunol.* **27**, 451–483 (2009).
- Biswas, S. K. & Mantovani, A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nature Immunol.* **11**, 889–896 (2010).
- Mantovani, A. & Sica, A. Macrophages, innate immunity and cancer: balance, tolerance, and diversity. *Curr. Opin. Immunol.* **22**, 231–237 (2010).
- Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
- Medzhitov, R. Origin and physiological roles of inflammation. *Nature* **454**, 428–435 (2008).

- Sica, A. & Mantovani, A. Macrophage plasticity and polarization: *in vivo* veritas. *J. Clin. Invest.* **122**, 787–795 (2012).
- Yokoyama, T. *et al.* *Trib1* links the MEK1/ERK pathway in myeloid leukemogenesis. *Blood* **116**, 2768–2775 (2010).
- Varbo, A., Benn, M., Tybjaerg-Hansen, A., Grande, P. & Nordestgaard, B. G. *TRIB1* and *GCKR* polymorphisms, lipid levels, and risk of ischemic heart disease in the general population. *Arterioscler. Thromb. Vasc. Biol.* **31**, 451–457 (2011).
- Aung, L. H. *et al.* Association of the *TRIB1* tribbles homolog 1 gene rs17321515 A>G polymorphism and serum lipid levels in the Mulao and Han populations. *Lipids Health Dis.* **10**, 230 (2011).
- Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature Genet.* **43**, 1131–1138 (2011).
- Yamamoto, M. *et al.* Enhanced TLR-mediated NF-IL6 dependent gene expression by *Trib1* deficiency. *J. Exp. Med.* **204**, 2233–2239 (2007).
- Jin, G. *et al.* *Trib1* and *Evi1* cooperate with *Hoxa* and *Meis1* in myeloid leukemogenesis. *Blood* **109**, 3998–4005 (2007).
- Keeshan, K. *et al.* *Tribbles* homolog 2 inactivates C/EBP α and causes acute myelogenous leukemia. *Cancer Cell* **10**, 401–411 (2006).
- Du, K., Herzig, S., Kulkarni, R. N. & Montminy, M. *TRB3*: A *tribbles* homolog that inhibits Akt/PKB activation by insulin in liver. *Science* **300**, 1574–1577 (2003).
- Okamoto, H. *et al.* Genetic deletion of *Trb3*, the mammalian *Drosophila* *tribbles* homolog, displays normal hepatic insulin signaling and glucose homeostasis. *Diabetes* **56**, 1350–1356 (2007).
- Kohyama, M. *et al.* Role for *Spi-C* in the development of red pulp macrophages and splenic iron homeostasis. *Nature* **457**, 318–321 (2009).
- De Domenico, I., McVey Ward, D. & Kaplan, J. Regulation of iron acquisition and storage: consequences for iron-linked disorders. *Nature Rev. Mol. Cell Biol.* **9**, 72–81 (2008).
- Hong, S., Skaist, A. M., Wheelan, S. J. & Friedman, A. D. AP-1 protein induction during monopoiesis favors C/EBP: AP-1 heterodimers over C/EBP homodimerization and stimulates *FosB* transcription. *J. Leukoc. Biol.* **90**, 643–651 (2011).
- Odegaard, J. I. & Chawla, A. Alternative macrophage activation and metabolism. *Annu. Rev. Pathol.* **6**, 275–297 (2011).
- Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J. Clin. Invest.* **117**, 175–184 (2007).
- Odegaard, J. I. *et al.* Macrophage-specific PPAR γ controls alternative activation and improves insulin resistance. *Nature* **447**, 1116–1120 (2007).
- Reese, T. A. *et al.* Chitin induces accumulation in tissue of innate immune cells associated with allergy. *Nature* **447**, 92–96 (2007).
- Satoh, T. *et al.* The *Jmjd3*-*Irf4* axis regulates M2 macrophage polarization and host responses against helminth infection. *Nature Immunol.* **11**, 936–944 (2010).
- Chawla, A., Nguyen, K. D. & Goh, Y. P. Macrophage-mediated inflammation in metabolic disease. *Nature Rev. Immunol.* **11**, 738–749 (2011).
- Gregor, M. F. & Hotamisligil, G. S. Inflammatory mechanisms in obesity. *Annu. Rev. Immunol.* **29**, 415–445 (2011).
- Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).
- Wellen, K. E. & Hotamisligil, G. S. Obesity-induced inflammatory changes in adipose tissue. *J. Clin. Invest.* **112**, 1785–1788 (2003).
- Wu, D. *et al.* Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science* **332**, 243–247 (2011).
- Odegaard, J. I. *et al.* Alternative M2 activation of Kupffer cells by PPAR δ ameliorates obesity-induced insulin resistance. *Cell Metab.* **7**, 496–507 (2008).
- Ricardo-Gonzalez, R. R. *et al.* IL-4/STAT6 immune axis regulates peripheral nutrient metabolism and insulin sensitivity. *Proc. Natl Acad. Sci. USA* **107**, 22617–22622 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank I. Shimomura and Y. Miyata for providing us with the key protocols for the metabolic experiments; T. Kitamura for providing the PlatE cells; M. Higa, H. Tanaka, N. Miyamoto, K. Miura, D. Ori, T. Uehata and K. Kuniyoshi for assistance with the experiments; and T. Kawai, S. Uematsu, T. Saitoh and Y. Kumagai for discussions. We also thank E. Kamada and M. Kageyama for secretarial assistance, and N. Umano, Y. Matsumoto and M. Kumagai for technical assistance. This work was supported by the Special Coordination Funds of the Japanese Ministry of Education, Culture, Sports, Science and Technology, and the Ministry of Health, Labour and Welfare in Japan, the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program).

Author Contributions T.S. designed and performed the experiments and wrote the manuscript. H.K., H.N. and N. Takakura performed the colony-forming assays. M.Y. generated the *Trib1*^{-/-}, *Trib2*^{-/-} and *Trib3*^{-/-} mice. N. Takemura helped with experiments. K.N. performed microarray analysis. Y.Y. performed the MRI experiments. E.M. performed the histological analyses. O.T. and S.A. designed experiments and wrote the manuscript. S.A. supervised the project.

Author Information Data have been deposited in the GEO under accession number GSE43563. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A. (sakira@biken.osaka-u.ac.jp).

METHODS

Mice, cells and reagents. *Trib1*^{-/-} mice were generated as described previously¹¹. Antibodies for flow cytometry were purchased from commercial sources as follows: anti-F4/80-FITC (BM8; BioLegend); anti-Mac1-PerCP Cy5.5 (M1/70; BD Biosciences); anti-Ly6C-FITC (AL-21; BD Biosciences); anti-Siglec-F-PE (E50-2446; BD Biosciences); anti-CCR3-APC (83103; BD Biosciences); anti-MR-Alexa 488 (MR5D3; BioLegend); and anti-CD11c-Brilliant bright (N418; BioLegend). Antibodies and reagents for histochemical analysis were purchased from commercial sources as follows: biotinylated anti-F4/80 (eBioscience), anti-MOMA-1 (Abcam), Alexa 488-conjugated anti-B220 (BD Pharmingen) and streptavidin-Alexa594 (Invitrogen). Antibodies for western blotting were purchased as follows: anti-C/EBP α (no. 2295; Cell Signaling Technology); anti-C/EBP β (no. E298; Millipore); anti-C/EBP δ (no. 2318; CST); anti-C/EBP ϵ (no. sc-158; Santa Cruz Biotechnology); and anti-actin (no. sc-1615; Santa Cruz Biotechnology). The probes for quantitative PCR were purchased from Lifescience Japan.

Generation of *Trib2*^{-/-} and *Trib3*^{-/-} mice. The *Trib2* and *Trib3* genes were isolated from genomic DNA extracted from embryonic stem (ES) cells (GSI-I) by PCR. The targeting vectors were constructed by replacing a 2-kilobase (kb) fragment encoding the *Trib2* open reading frame (exon 2) and replacing a 2.5-kb fragment encoding the *Trib3* open reading frame (exons 2–3) with a neomycin-resistance gene cassette (*neo*), and inserting herpes simplex virus thymidine kinase (HSV-TK) driven by the PGK promoter into the genomic fragment for negative selection. After each targeting vector was transfected into ES cells, G418 and ganciclovir double-resistant colonies were selected and screened by PCR, and recombination was further confirmed by Southern blotting. The homologous recombinant clones were individually microinjected into blastocysts derived from C57BL/6 mice and then transferred to pseudopregnant females. Mating of chimaeric male mice with C57BL/6 female mice resulted in the transmission of each mutant allele to the germ line. The resulting *Trib2*^{+/-} and *Trib3*^{+/-} mice were interbred to generate *Trib2*^{-/-} and *Trib3*^{-/-} mice, respectively. All animal experiments were performed with approval from the Animal Research Committee of the Research Institute for Microbial Diseases (Osaka University).

Construction of expression plasmids and shRNA vectors. *Trib1* and *Cebpa* cDNAs were obtained by PCR from a mouse cDNA library. Full-length or mutated *Trib1* cDNAs were cloned into the pLZR-cmv-ires-GFP vector for retrovirus production. For shRNA experiments, the following oligonucleotides were used to target mouse C/EBP α : *cebpa*-8 dn, 5'-gatccGTCGTATGTTATATCTATATTCACAGATATAGATATAATACATACGACTTTTTCACGCGTg-3' and *cebpa*-8 up, 5'-aattcACGCGTAAAAAGTCGTATGTTATATATCTATATCTCTGAATATAGATATAATACATACGACg-3'; *cebpa*-12 dn, 5'-gatccATCCGATATCAACACTTGTATTTCACAGAGAATACAAGTGTGATATCGGATT TTTTACGCGTg-3' and *cebpa*-12 up, 5'-aattcACGCGTAAAAAATCCGATTCACACTTGTATTCTCTTGAATACAAAGTGTGATATCGGATg-3'. The oligonucleotides were annealed and ligated with the RNAi-ready pSIREN-RetroQ-zsGreen vector (Takara). Lowercase letters indicate the site of restriction enzyme.

Colony-forming assay. Bone marrow cells (4×10^4) were seeded on 3.5-cm dishes and cultured for 7 days with Methocult (no. 3534; Stem Cell Technologies) containing SCF, IL-6 and IL-3 in the presence of 20 ng ml^{-1} M-CSF. The numbers of GFP-positive cells were counted for colony-forming assays or collected for real-time (q)PCR and western blotting.

Assessment of glucose, insulin and NEFA levels and glucose and insulin sensitivity *in vivo*. Serum insulin levels were determined by ELISA (Moringa Institute of Biological Science Inc.). Serum glucose, NEFA and cholesterol levels were measured with enzymatic kits (WAKO Chemicals). Intraperitoneal glucose tolerance tests were performed on 16 h fasted mice injected intraperitoneally with D-glucose (1.5 mg g^{-1} body weight). Serum glucose levels were measured immediately before and 30, 60, 90 and 120 min after the D-glucose injection. Intraperitoneal insulin tolerance tests were performed on 4 h fasted mice injected intraperitoneally with insulin (1 mU g^{-1} body weight). Serum glucose levels were measured immediately before and 15, 30, 60 and 90 min after the insulin injection.

Lipolysis assay. *In vivo* lipolysis assays were performed on 4 h fasted mice. Blood was collected from orbital veins, and the serum glycerol and NEFA concentrations were measured.

Quantitative PCR analysis. Total RNA was isolated using Trizol (Invitrogen) and subjected to reverse transcription with ReverTra Ace (Toyobo) according to the manufacturer's instructions. For quantitative PCR, cDNA fragments were amplified using real-time PCR Master Mix (Toyobo). Fluorescence from the TaqMan probe for each cytokine was detected using a 7500 Real-time PCR System (Applied Biosystems). To determine the relative induction of each cytokine mRNA in response to various stimuli, the mRNA expression level of each gene was normalized to the 18S rRNA expression level. The experiments were repeated at least twice.

Assay ID numbers are: *Arg1*, Mm00475988_m1, *Ym1*, Mm00657889_mH, *Fizz1*, Mm00445109_m1, *MR*, Mm00485148_m1; *Trib1*, Mm00454875_m1; *Trib2*, Mm00045 4876_m1; *Trib3*, Mm00454879_m1; *Acaca*, Mm01304277_m1; *Fasn*, Mm00662319_m1; *Acs1l*, Mm00484217_m1; *PParg*, Mm01184322_m1; *Cebpa*, Mm00514283_s1; *Cebpb*, Mm00843434_s1; *Cebpd*, Mm00786711_s1; *Il10*, Mm00439614_m1; *Tnfa*, Mm99999068_m1; *Inos*, Mm00440485_m1.

Immunoblot analysis. Bone marrow cells were collected and red blood cells were lysed using red blood cell lysis buffer (Invitrogen). After two washes with MACS buffer, lineage-negative cells were collected by MACS sorting using anti-CD3, anti-CD19, anti-B220, anti-Ter119, anti-Gr1, anti-DX5, anti-Mac1 and anti-CD11c MACS beads (Miltenyi Biotec). The lineage-negative cells were lysed with lysis buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP40) containing Complete Protease Inhibitor Cocktail (Roche). The cell lysates were separated by standard SDS-PAGE and analysed by immunoblotting. Antibodies against the following proteins were used: C/EBP α , C/EBP β , C/EBP δ , C/EBP ϵ and actin. The Luminata Forte Western HRP Substrate (Millipore) was used for the development of positive signals.

Flow cytometry. Cell suspensions were prepared by sieving and gentle pipetting. After washing with ice-cold FACS buffer (0.5% BSA and 2 mM EDTA in PBS, pH 7.2), the cells were incubated with antibodies for 15 min and washed twice with FACS buffer. Data were acquired using a FACSCalibur flow cytometer (BD), and analysed by FlowJo software (Tree Star).

Retroviral preparation. Retroviruses were produced using PlatE packaging cells or transfected with various plasmids. Bone marrow cells were transduced with the retroviral supernatants (supplemented with polybrene (Millipore)) twice. After the second transduction, the cells were washed twice and re-suspended in RPMI medium.

Magnetic resonance imaging (MRI). An 11.7 T MRI scanner (AVANCE-II 500 WB; Bruker BioSpin) was used to acquire *in vivo* mouse images. Mice were anaesthetized with 1.0–1.5% isoflurane during the MRI procedures. We used a T₂-weighted imaging sequence (RARE) with the following parameters: field of view, 2.5×2.5 cm; matrix size, 256×256 ; slice thickness, 0.5 mm; repetition time, 5000 ms; echo time, 25.2 ms; average, 4; scan time, 5 min.

Histochemical analysis. Frozen 6- μm tissue sections were fixed in 4% paraformaldehyde and blocked with 5% BSA (Invitrogen). For Perl's Prussian blue staining, tissues were fixed with 4% paraformaldehyde in phosphate buffer (pH 7.0), embedded in paraffin and stained with Perl's Prussian blue.

Generation of bone-marrow-transferred chimaeric mice. Bone marrow cells were prepared from wild-type and *Trib1*^{-/-} mice, and intravenously injected into lethally irradiated CD45.1 C57BL/6 mice, wild-type and *Trib1*^{-/-} mice. The chimaeric mice were administered neomycin (Sigma) and polymyxin B (Sigma) in their drinking water for 8 weeks. The mice were analysed for at least 12 weeks after reconstitution. More than 90% of splenocytes from the chimaeric mice were CD45.2⁺.

Mouse SVF and MAF isolation. Epididymal fat pads from male mice fed a normal or HFD were excised and minced in Krebs-Ringer-bicarbonate-HEPES buffer (KRBH: 120 mM NaCl, 4 mM KH₂PO₄, 1 mM MgSO₄, 1 mM CaCl₂, 10 mM NaHCO₃, 30 mM HEPES, 20 μM adenosine (Sigma), 4% BSA). The tissues were then digested with 1 mg ml^{-1} collagenase type II (Sigma) in PBS at 37 °C for 30 min with gentle shaking and filtered through a 100- μm filter. In SVF preparations, the digested cells were separated by centrifugation at 700g for 5 min. In MAF preparations, the digested cells were separated by centrifugation at 300g for 5 min. The resulting cells were washed twice with KRBH buffer before use in experiments.

Metabolic measurements. Serum insulin levels were determined by ELISA (Moringa Institute of Biological Science Inc.). Serum glucose, NEFA and cholesterol levels were measured with enzymatic kits (WAKO Chemicals). Intraperitoneal glucose tolerance tests were performed on 16 h fasted mice injected intraperitoneally with D-glucose (1.5 mg g^{-1} body weight). Serum glucose levels were measured immediately before administration and 30, 60, 90 and 120 min after the D-glucose injection.

Lipolysis assay. *In vivo* lipolysis assays were performed on 4 h fasted mice. Blood was collected from orbital veins, and the serum glycerol and NEFA concentrations were measured. For *ex vivo* assays, adipocytes were isolated from epididymal pads using collagenase, washed twice, re-suspended in Krebs-Ringer buffer at 5×10^5 cells ml^{-1} and incubated with shaking for the specified times. The glycerol and NEFA concentrations were measured in the supernatants.

Microarray analysis. Total RNA was isolated from wild-type and *Trib1*^{-/-} GMP and MDP using TRIzol RNA isolation kit (Invitrogen) and further purified using an RNeasy kit (Qiagen). Biotinylated cDNA was synthesized from 100 ng total RNA with the Ovation biotin RNA amplification and labelling systems (Nugen) according to the manufacturer's protocol. The product was purified using an DyeEx 2.0 spin kit (QIAGEN), fragmented, and hybridized to Affymetrix mouse

expression array A430 2.0 microarray chips, according to the manufacturer's protocol (Affymetrix). Staining, washing and scanning of Affymetrix mouse Genome 430 2.0 microarray chips was done following the manufacturer's instructions. Robust multichip average (RMA) expression values were calculated using R and the Bioconductor affy package. For each probe the changes in expression between wild-type and *Trib1*^{-/-} samples were defined as the difference between

log₂ values for wild-type and *Trib1*^{-/-} cells. Genes were assigned the values of their corresponding probe(s). In cases where a gene was associated with multiple probes the average value was taken.

Statistical analysis. The statistical significance of differences between values was calculated using the two-tailed Student's *t*-test. Values of **P* < 0.05, ***P* < 0.01 were considered to indicate statistical significance.

CLASP-mediated cortical microtubule organization guides PIN polarization axis

Klementina Kakar¹, Hongtao Zhang^{1,2,3}, Ben Scheres^{1,3,4} & Pankaj Dhonukshe^{1,5,6}

Recent evidence indicates a correlation between orientation of the plant cortical microtubule cytoskeleton and localization of polar cargoes¹. However, the molecules and mechanisms that create this correlation have remained unknown. Here we show that, in *Arabidopsis thaliana*, the microtubule orientation regulators CLASP^{2,3} and MAP65 (refs 3, 4) control the abundance of polarity regulator PINOID kinase^{5,6} at the plasma membrane. By localized upregulation of clathrin-dependent endocytosis⁷ at cortical microtubule- and clathrin-rich domains orthogonal to the axis of polarity, PINOID accelerates the removal of auxin transporter PIN proteins^{8–10} from those sites. This mechanism links directional microtubule organization to the polar localization of auxin transporter PIN proteins, and clarifies how microtubule-enriched cell sides are kept distinct from polar delivery domains. Our results identify the molecular machinery that connects microtubule organization to the regulation of the axis of PIN polarization.

In animals, the cytoskeleton establishes and sustains polar domains¹¹. In some plant cells, actin and microtubule networks form mutually exclusive structures regulated by plant Rho homologues¹², mediating polarization of PIN family auxin efflux carriers^{13,14}. Notably, cortical microtubule orientation correlates with PIN polarization in the shoot apex¹, suggesting another link between cytoskeletal configurations and PIN localization. Auxin is a key plant signalling molecule with numerous roles in development^{8–10}. Polar localization of PIN proteins regulates cell-to-cell auxin transport and auxin distribution-dependent plant patterning^{8–10}. Therefore, mechanisms regulating PIN localization are crucial for plant morphogenesis. Regulation of PIN phosphorylation through PINOID (PID) kinases restricts PIN trafficking to one of two opposing polar domains^{5,6}, but how the axis of polarization is selected remains unsolved. Microtubule regulators CLASP and MAP65 are required for precise switches in cortical microtubule orientation in plant cells³. Reasoning that, like in animals, plant microtubules might control the cell polarization axis and that microtubule organization and PIN polarization correlate¹, we investigated whether CLASP regulates the axis of PIN polarity.

PIN2::PIN2-GFP in wild-type cells reveals apical ('shootward') localization in the epidermis, and basal ('rootward') localization in the cortex⁸ (Fig. 1a). In *clasp-1* mutant cells¹⁵, PIN2 localization shifted from rootwards to shootwards in the cortex, with reduced PIN2 levels compared to control (Fig. 1a–c and Supplementary Fig. 1). Polar localization of PIN2 became sharper in *clasp-1* mutants, without any alterations in PIN2 transcript levels (Fig. 1d–f; see Methods and Supplementary Fig. 2 for details on intensity quantifications). Induction of CLASP in *clasp-1* mutants largely rescued cortical PIN2 localization defects (Supplementary Fig. 3a–g). Rootward-to-shootward shifts in PIN2 localization occur in elongating cortical cells of the wild type, but not in the root meristematic cortical cells of many short root meristem mutants, indicating that the PIN2 switch in the *clasp-1* cortex was not related to its reduced root meristem (Supplementary

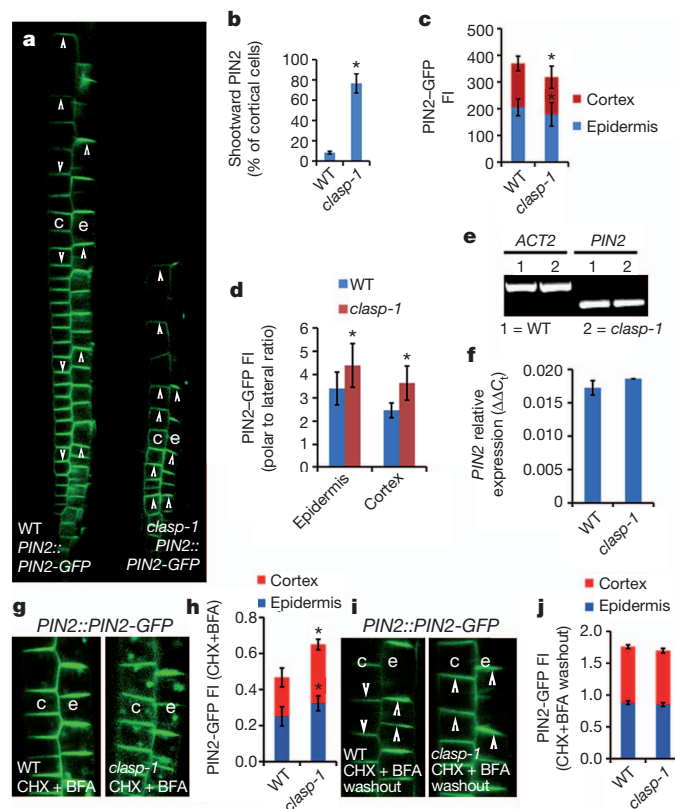


Figure 1 | PIN2 localization and trafficking is altered in the *clasp-1* mutant. a–c, PIN2::PIN2-GFP localization in wild type (WT) and *clasp-1* mutants ($n > 16$ roots) (a, b), and PIN2-green fluorescent protein (GFP) intensity at the plasma membrane in the epidermis and cortex of wild type ($n > 172$ cells from 18 roots) and *clasp-1* mutants ($n > 188$ cells from 20 roots) (c). FI, fluorescence intensity. d, Polar to lateral PIN2-GFP intensity ratio in epidermis and cortex of wild type ($n > 49$ cells from 8 roots) and *clasp-1* mutants ($n > 56$ cells from 10 roots). e, f, Expression of ACTIN 2 (ACT2) and PIN2 in wild type and *clasp-1* mutants as detected by PCR with reverse transcription (RT-PCR) (e), and quantitative RT-PCR (qRT-PCR) (f). g–j, PIN2-GFP aggregation into BFA compartments in the epidermis and cortex of wild type and *clasp-1* mutants after 30 min CHX pre-treatment followed by 30 min BFA plus CHX treatment (g, h); and PIN2 recycling to the plasma membrane in epidermis and cortex of wild type and *clasp-1* after 1 h BFA washout in the presence of CHX (i, j). h, PIN2-GFP endocytosis in wild type ($n > 65$ cells from 8 roots) and *clasp-1* mutants ($n > 52$ cells from 8 roots). j, PIN2-GFP recycling in wild type ($n > 82$ cells from 9 roots) and *clasp-1* mutants ($n > 40$ cells from 8 roots). c, cortex; e, epidermis. Green denotes GFP. Arrowheads depict PIN2 polarity. Data are mean and s.d. * $P < 0.001$ (Student's *t*-test). Original magnification for all confocal images, $\times 63$.

¹Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. ²Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands. ³Netherlands Proteomics Centre, Padualaan 8, 3584CH Utrecht, The Netherlands. ⁴Wageningen University Research, Droeveendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁵Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium. ⁶Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium.

Fig. 3l). Rootward PIN1 localization in vasculature and its transcript levels remained unaffected in *clasp-1* mutants (Supplementary Fig. 4a–h). Moreover, ectopically expressed PIN2 remained rootward in vasculature of the wild type and *clasp-1* mutant (Supplementary Fig. 6a, c). Our data indicate that CLASP regulates PIN2 localization in the root cortex.

Treatment with recycling inhibitor brefeldin A (BFA) and translation inhibitor cyclohexamide (CHX)¹⁶ resulted in increased PIN2 internalization into BFA compartments in epidermal and cortical cells of *clasp-1* as compared to cells of wild type (Fig. 1g, h and Supplementary Fig. 3m, n). BFA washout in the presence of CHX restored PIN2 localization at the plasma membrane in wild type and *clasp-1* cells, suggesting intact PIN2 recycling in *clasp-1* mutants (Fig. 1i, j and Supplementary Fig. 3o, p). Endocytosis and recycling of *PIN1::PIN1-GFP* remained unaffected in *clasp-1* mutants (Supplementary Fig. 4i–n). Our results reveal enhanced PIN2 endocytosis, and operational PIN2 recycling, in the *clasp-1* cortical and epidermal layers.

Because PIN2-mediated auxin distribution is implicated in the gravitational response⁹ and root meristem patterning⁸, and PIN2 localization and trafficking are altered in *clasp-1* mutants, we analysed auxin-related root phenotypes. *clasp-1* mutant roots showed reduced growth, reduced meristem size, cellular organization defects, abnormal gravitropic growth, and delayed responses to stimulation by gravity (gravistimulation; Supplementary Fig. 5a–h). Induction of CLASP in *clasp-1* mutants largely rescued these defects (Supplementary Fig. 3a–c, h–k). Although gravistimulated wild-type roots displayed a strong asymmetric distribution of the auxin activity reporter *DR5::YFP-NLS*³, this response was delayed and impaired in *clasp-1* roots (Supplementary Fig. 5i, j). Furthermore, application of the auxin analogues indole-3-acetic acid (IAA), 1-naphthaleneacetic acid (NAA), auxin efflux inhibitor 1-N-naphthylphthalamic acid (NPA), and endogenous auxin

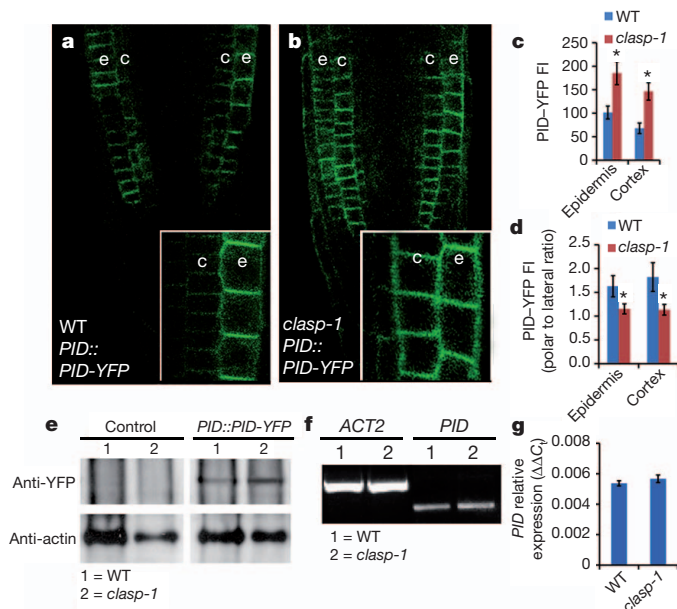


Figure 2 | PINOID abundance is enhanced in the *clasp-1* mutant. **a–c**, *PID::YFP* localization (**a**, **b**) and fluorescence intensity (**c**) at the plasma membrane in epidermis and cortex of wild type ($n > 28$ cells from 8 roots) and *clasp-1* mutants ($n > 26$ cells from 7 roots). **d**, Polar to lateral *PID::YFP* intensity ratio in epidermis and cortex of wild type ($n > 38$ cells from 10 roots) and *clasp-1* mutants ($n > 30$ cells from 8 roots). **e**, Total protein fractions of wild type, *clasp-1*, *PID::YFP* and *clasp-1* + *PID::YFP* seedlings probed with anti-YFP and anti-actin antibodies. **f**, **g**, Expression of *ACTIN 2* and *PID* in wild type and *clasp-1* mutants as detected by RT-PCR (**f**) and qRT-PCR (**g**). Green denotes YFP. Data are mean and s.d. * $P < 0.001$ (Student's *t*-test). Original magnification for all confocal images, $\times 63$.

overproduction³ enhanced *DR5::YFP-NLS* expression (Supplementary Fig. 5k–m) and partially rescued cellular organization defects of *clasp-1* roots (Supplementary Fig. 5r, s), but not meristem size (Supplementary Fig. 5n–q). This suggests functional auxin perception and responses in *clasp-1* mutants, and the causal relation of *clasp-1* root meristem defects with shootward PIN2-mediated changes in auxin transport. In accordance, expression of rootward *PIN2::PIN1-GFP2* (ref. 9) in *clasp-1* mutants partially resumed rootward auxin flux and slightly rescued *clasp-1* root defects (Supplementary Fig. 5t–w).

The decision for shootward or rootward PIN localization depends on PID action^{5,6}. In wild-type roots, high epidermal plasma membrane-bound PID levels efficiently phosphorylate PIN2 for its shootward translocation, whereas low cortical plasma membrane-bound PID levels inefficiently phosphorylate PIN2 for its rootward localization⁶ (Fig. 2a). *PID::YFP* displayed more intense plasma membrane localization signals and higher protein levels in *clasp-1* mutants compared to wild type (Fig. 2b, c, e), without changes in transcript levels (Fig. 2f, g). CLASP induction in *clasp-1* mutants restored PID abundance (Supplementary Fig. 6o, p), and PID induction switched PIN2 from rootward to shootward cell sides in the vasculature and cortex (Supplementary Fig. 6a–d), implicating increased PID levels in *clasp-1* mutants for PIN2 relocation. Accordingly, cortically induced *PID::YFP* (*CO2::XVE* \gg *PINOID-YFP*) triggered rootward-to-shootward PIN2 shifts (Supplementary Fig. 6e, f), and mimicked organ level phenotypes observed in the *clasp-1* mutant (Supplementary Fig. 6g–k),

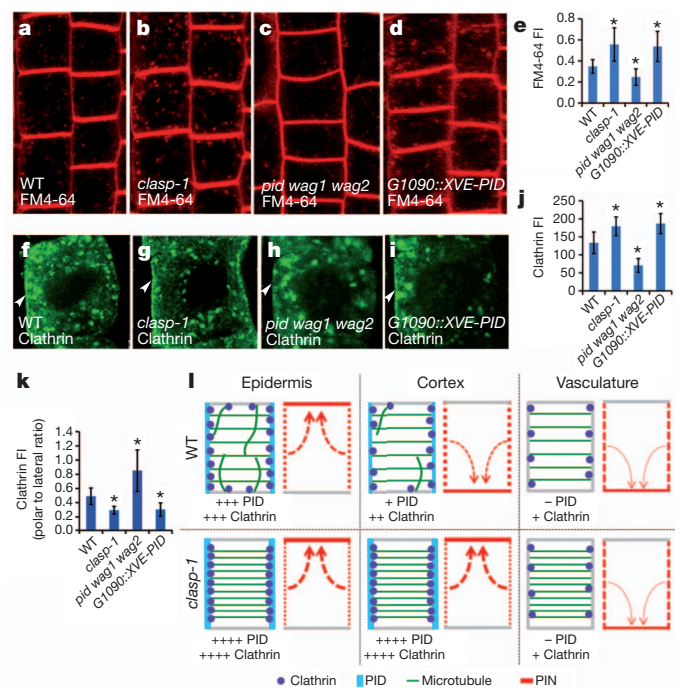


Figure 3 | Endocytosis is enhanced in *clasp-1* mutants by a PID-dependent increment in plasma-membrane localized clathrin. **a–e**, Evaluation of endocytosis in wild type (**a**), *clasp-1* (**b**), *pid wag1 wag2* (**c**) and *G1090::XVE-PID* (**d**) roots by analysing FM4-64 internalization (**e**) ($n > 45$ cells from 10 roots each). **f–i**, Clathrin localization in wild type (**f**), *clasp-1* (**g**), *pid wag1 wag2* (**h**) and *G1090::XVE-PID* (**i**) roots. **j**, **k**, Clathrin intensity at the plasma membrane ($n > 24$ cells from 7 roots each) (**j**) and its polar to lateral intensity ratio ($n > 44$ cells from 9 roots each) (**k**). **l**, Schematic illustrating the correlation between the microtubule orientation-related levels of lateral PIN and lateral clathrin to the amount of lateral PIN removal in epidermal, cortical and vascular cells of wild type and *clasp-1* mutants. Line and arrow thicknesses, as well as '+' and '-', depict relative amounts of PID, clathrin and PIN. Red denotes FM4-64, green denotes anti-CHC (clathrin heavy chain). Arrowheads highlight CHC localization at the plasma membrane. Data are mean and s.d. * $P < 0.001$ (Student's *t*-test). Original magnification for all confocal images, $\times 63$.

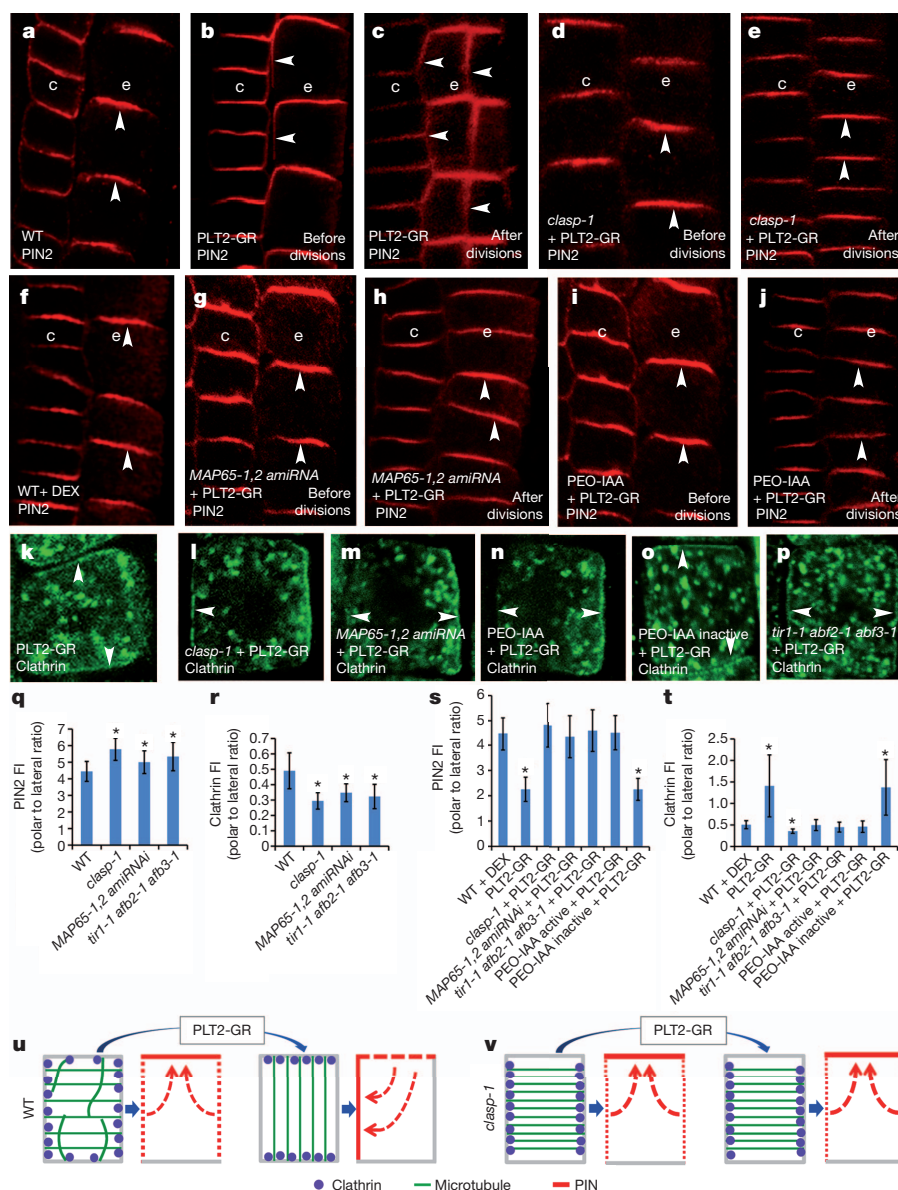


Figure 4 | Microtubule orientation defines the axis of PIN polarization. **a, b**, PIN2 localization in wild type (rootward in cortex; shootward in epidermis) (**a**) and after PLT2-GR induction (shootward–lateral in epidermis) (**b**). **c**, PLT2 induced shootward–lateral epidermal PIN2 localization persists after periclinal cell divisions. **d–j** PIN2 polarity after PLT2-GR induction in *clasp-1* mutants (**d, e**), 35S::MAP65-1,2 *amiRNA* (**g, h**) and after PEO-IAA treatment (largely shootward in cortex and epidermis) (**i, j**). Normal PIN2 localization in wild type after DEX treatment (**f**). **k–p**, After PLT2-GR induction, clathrin localizes at polar cell sides in the epidermis of wild type (**k**) and after inactive PEO-IAA treatment (**o**), and at lateral cell sides in epidermis of *clasp-1* mutants (**l**), 35S::MAP65-1,2 *amiRNA* mutants (**m**), after active PEO-IAA treatment (**n**), and in *tir1-1 afb2-1 afb3-1* mutants (**p**). PLT2-GR induction followed by inactive PEO-IAA treatment results in clathrin localization at the polar cell sides of epidermis (**o**). **q, s**, Polar to lateral PIN2

intensity ratio ($n > 36$ cells from 8 roots each (**q**); $n > 27$ cells from 8 roots each (**s**)). **r, t**, Polar to lateral clathrin intensity ratio ($n > 27$ cells from 8 roots each (**r**); $n > 19$ cells from 8 roots each (**t**)). **u, v**, Schematic illustrating microtubule array organization and PIN2 polarity in epidermal cells of wild type (**u**) and *clasp-1* mutants (**v**), following PLT2-GR induction. Note, in wild type, *clasp-1* mutants and *clasp-1* mutants plus PLT2-GR induction, PIN2 localizes to the polar cell side, aligning with transversal microtubule array. After PLT2-GR induction in the wild type, the microtubules reorient from transversal to longitudinal, and accordingly PIN2 shifts its localization from the polar to lateral cell side. Line and arrow thicknesses depict the amounts. In **a–p**, red denotes PIN2, green denotes CHC; arrowheads depict PIN2 or CHC localizations. 35S::MAP65-1,2 *amiRNA* targets MAP65-1 and MAP65-2 (ref. 3). Data are mean and s.d. * $P < 0.001$ (Student's *t*-test). Original magnification for all confocal images, $\times 63$.

whereas PID repression largely rescued PIN2 defects in *clasp-1* mutants (Supplementary Fig. 6l–n). Together, our results indicate that loss of CLASP increases PID levels at the plasma membrane to influence PIN2 localization.

Because increased PIN2 endocytosis correlated with enhanced PID levels in *clasp-1* mutants, we tested whether PID affected endocytosis. We probed endocytosis by internalization of the endocytic tracer FM4-64 (ref. 17). Fewer FM4-64-labelled endocytic vesicles were present in *pid wag1 wag2* (ref. 6) triple mutants, and more were present

in *clasp-1* mutants and in induced *G1090::XVE* \gg *PID*, than in wild type (Fig. 3a–e), indicating the influence of PID on endocytosis. As clathrin is a major effector of endocytosis in plants⁷ we tested clathrin localization. Clathrin levels increased at the *clasp-1* mutant plasma membrane and after PID induction, particularly at the lateral domains, whereas they decreased at the *pid wag1 wag2* mutant plasma membrane, as compared to wild type (Fig. 3f–k), indicating a positive correlation of PID to clathrin levels to upregulate PIN endocytosis (Fig. 3l). In accordance, PID-induction-triggered PIN relocation occurs rapidly

(by upregulation of PIN endocytosis)⁶, and the stress-induced shift in PIN relocation (requiring PIN endocytosis) is impaired in a *pid* mutant¹.

Owing to impaired microtubule crossover from lateral to top–bottom polar cell sides^{2,3}, lateral microtubules were more predominant in the *clasp-1* cortex and epidermis (Supplementary Fig. 7a–e). This skewed microtubule distribution enhanced PID levels at the lateral cell sides of epidermal and cortical cell layers in *clasp-1* mutants (Fig. 2d and Supplementary Fig. 7h), and CLASP induction in *clasp-1* mutants restored this defect (Supplementary Fig. 6p), suggesting a causal link between axial microtubule density and PID enrichment. Oryzalin treatment, which depolymerizes microtubules and detaches CLASP from microtubules and cell periphery³, reduced lateral PID levels, decreased clathrin abundance and internalization of FM4-64 (ref. 17; Supplementary Fig. 7f–n), and increased PIN2 lateral abundance (Supplementary Fig. 7o–q). Microtubular manipulation suggests that CLASP-dependent microtubule organization regulates lateral PID, clathrin and endocytosis levels, which in turn control lateral PIN removal.

CLASP regulates microtubule reorientation in conjunction with PLETHORA (PLT) transcription factors, which induce MAP65 microtubule regulators through TIR1 auxin signalling³. MAP65-1 and MAP65-2 activity are required for CLASP function³. Notably, roots of MAP65-1,2 *amiRNA* lines (which target MAP65-1 and MAP65-2)³, *map65-1 map65-2* double mutants³ and *tir1-1 afb2-1 afb3-2* auxin signalling triple mutants¹⁸ displayed root gravitropism, meristem size and PIN2 localization defects, all reminiscent of *clasp-1* phenotypes (Supplementary Fig. 8a–k). Although PID messenger RNA levels remained unaffected, the auxin signalling antagonist PEO-IAA¹⁹ and triple-mutant *tir1-1 afb2-1 afb3-1* increased PID levels at the lateral plasma membrane (Supplementary Fig. 8l–r), indicating the requirement of TIR1 and MAP65 for CLASP-mediated PID localization and, consequently, for defining the axis of PIN2 localization.

To test directly whether TIR1 and MAP65 activation influences the polarization axis, we used the dexamethasone (DEX)-inducible PLT2-GR system³, which triggers cortical microtubule reorientation³. At ~18 h after PLT induction, PIN2 localization shifted from shootward towards lateral cell sides in root epidermal cells of wild type but not of *clasp-1* mutants, correlating with microtubule crossover at top–bottom cell sides (Fig. 4a, b, d, f and Supplementary Fig. 7a–e), without altering PIN2 transcript levels (Supplementary Fig. 10k). This suggests that the axis of PIN2 localization in root epidermal cells is modified after microtubule reorientation but before cell division plane rotation (Fig. 4a–e, s). This PLT2 effect was suppressed in MAP65-1,2 *amiRNA* (Fig. 4g, h, s), in *tir1-1 afb2-1 afb3-1* seedlings, and in seedlings treated with PEO-IAA (Fig. 4i, j, s). Together, these results indicate that CLASP, and TIR1 auxin signalling-induced MAP65 action can define a microtubule-dependent axis of PIN polarity.

PLT2-GR induction and NPA treatment remained able to reorient microtubules in mutants for multiple PINs⁸, and in *pid wag1 wag2* mutants (Supplementary Fig. 10a–j), indicating PIN- and PID-independent microtubule reorientation, in line with previous work¹.

Shootward epidermal PIN2 localization increased (Fig. 4q) and clathrin levels heightened at lateral cell sides compared to shootward and rootward cell sides in root epidermal cells of *clasp-1*, 35S::MAP65-1,2 *amiRNA* and *tir1-1 afb2-1 afb3-1* mutants (Fig. 4r), whereas clathrin transcript levels were unaltered (Supplementary Fig. 9h). These data indicated more PIN2 removal from lateral cell sides in those mutants than in wild type. PLT2-GR induction and co-treatment with inactive PEO-IAA in wild type shifted the predominant localization of clathrin heavy and light chains from the lateral to shootward–rootward cell sides of root epidermal cells within the 16 h time frame that reorients microtubules (Fig. 4k, o, t and Supplementary Fig. 9d, f, g). By contrast, predominant lateral localization of clathrin in root epidermal cells remained unaltered after PLT2 induction in *clasp-1*, in 35S::MAP65-1,2 *amiRNA*, in *tir1-1 afb2-1 afb3-1* mutant backgrounds, and in wild type after active PEO-IAA co-treatment, where PLT2 induction fails to reorient microtubules (Fig. 4l–n, p, t and Supplementary Fig. 9a–c, e, g). Together, these findings indicate that

microtubule reorientation switches the axis of enhanced microtubule and clathrin localization, which in turn contributes to the switch in PIN2 localization by changing the axial bias of PIN2 endocytosis (Fig. 4u, v).

Our findings connect CLASP- and MAP65-mediated microtubule array orientation to the establishment of the polarity axis in plant cells. The data suggest a mechanism in which CLASP controls microtubule organization over different surfaces. Given the enhanced microtubule co-alignments in cell faces perpendicular to the axis of polarity in *clasp-1* mutant roots², we predict a mechanism in which high transversal microtubule organization increases PID and clathrin levels, which locally stimulate PIN endocytosis. In accordance, enhanced clathrin localization and enhanced endocytosis have been shown to occur at the oriented microtubule dense regions during preprophase stage^{20,21}. Differential endocytosis then favours PIN retention at the cell faces through which the axis of polarity passes.

Our finding unravels the inner workings of the plant cell polarization machinery by pinpointing a mechanism to restrict the cellular polarity axis. Furthermore, PID now seems to have two functions: defining the axis of polarity, and defining polarity itself within this polar axis^{5,6}. Our data indicate that these functions (1) depend on the plasma membrane PID levels, which influence the axis of polarity by promoting endocytosis along axial cell faces, and (2) also control the polar recycling to one of two opposing cell faces by their absolute levels and phosphorylation activity. How the polarity within this axis is determined, how PID proteins perform both endocytosis and polar targeting functions, and whether PID action on clathrin-mediated endocytosis is locally regulated remains to be established.

METHODS SUMMARY

Plant material and growth conditions. All *Arabidopsis* lines are detailed in Methods. Seedlings were grown as described previously¹⁰. All experiments represent at least three independent biological replicates.

PIN polarity. For PIN polarity quantification, shootward PIN2 percentages were quantified by counting the cells with shootward PIN2 and expressing these compared to the total number of cortical cells in the root apical meristem.

Chemical treatments. IAA (Ducheva), NAA (Ducheva), NPA (Ducheva), IAM (Ducheva), BFA (Sigma), CHX (Sigma), oestradiol (Sigma), oryzalin (Sigma), PEO-IAA (a gift from K. Hayashi) and DEX (Sigma) were used from dimethylsulphoxide-dissolved stock solutions as indicated.

Immunolocalization. Whole-mount immunolocalization was performed as described previously¹⁰.

Confocal image analysis. Fluorescence signal intensity was measured using the Leica (live) confocal software in a region of interest. Polar and lateral fluorescence intensity values were obtained by manually drawing lines on lateral and polar cell sides. Their ratios were obtained by averaging polar to lateral fluorescence intensity ratios of individual cells.

Full Methods and any associated references are available in the online version of the paper.

Received 8 October 2012; accepted 31 January 2013.

Published online 20 March 2013.

1. Heisler, M. G. *et al.* Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. *PLoS Biol.* **8**, e1000516 (2010).
2. Ambrose, C., Allard, J. F., Cytrynbaum, E. N. & Wasteneys, G. O. A CLASP-modulated cell edge barrier mechanism drives cell-wide cortical microtubule organization in *Arabidopsis*. *Nature Commun.* **2**, 430 (2011).
3. Friml, J. *et al.* A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science* **306**, 862–865 (2004).
4. Dhonukshe, P. *et al.* A PLETHORA-auxin transcription module controls cell division plane rotation through MAP65 and CLASP. *Cell* **149**, 383–396 (2012).
5. Smertenko, A. *et al.* A new class of microtubule-associated proteins in plants. *Nature Cell Biol.* **2**, 750–753 (2000).
6. Friml, J. *et al.* A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science* **306**, 862–865 (2004).
7. Dhonukshe, P. *et al.* Plasma membrane-bound AGC3 kinases phosphorylate PIN auxin carriers at TPRXS(N/S) motifs to direct apical PIN recycling. *Development* **137**, 3245–3255 (2010).
8. Dhonukshe, P. *et al.* Clathrin-mediated constitutive endocytosis of PIN auxin efflux carriers in *Arabidopsis*. *Curr. Biol.* **17**, 520–527 (2007).
9. Blilou, I. *et al.* The PIN auxin efflux facilitator network controls growth and patterning in *Arabidopsis* roots. *Nature* **433**, 39–44 (2005).

9. Wisniewska, J. *et al.* Polar PIN localization directs auxin flow in plants. *Science* **312**, 883 (2006).
10. Dhonukshe, P. *et al.* Generation of cell polarity in plants links endocytosis, auxin distribution and cell fate decisions. *Nature* **456**, 962–966 (2008).
11. Li, R. & Gundersen, G. G. Beyond polymer polarity: how the cytoskeleton builds a polarized cell. *Nature Rev. Mol. Cell Biol.* **9**, 860–873 (2008).
12. Fu, Y., Gu, Y., Zheng, Z., Wasteneys, G. & Yang, Z. *Arabidopsis* interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. *Cell* **120**, 687–700 (2005).
13. Nagawa, S. *et al.* ROP GTPase-dependent actin microfilaments promote PIN1 polarization by localized inhibition of clathrin-dependent endocytosis. *PLoS Biol.* **10**, e1001299 (2012).
14. Xu, T. *et al.* Cell surface- and rho GTPase-based auxin signaling controls cellular interdigitation in *Arabidopsis*. *Cell* **143**, 99–110 (2010).
15. Ambrose, J. C., Shoji, T., Kotzer, A. M., Pighin, J. A. & Wasteneys, G. O. The *Arabidopsis* CLASP gene encodes a microtubule-associated protein involved in cell expansion and division. *Plant Cell* **19**, 2763–2775 (2007).
16. Geldner, N., Friml, J., Stierhof, Y. D., Jurgens, G. & Palme, K. Auxin transport inhibitors block PIN1 cycling and vesicle trafficking. *Nature* **413**, 425–428 (2001).
17. Dhonukshe, P. *et al.* Endocytosis of cell surface material mediates cell plate formation during plant cytokinesis. *Dev. Cell* **10**, 137–150 (2006).
18. Dharmasiri, N. *et al.* Plant development is regulated by a family of auxin receptor F box proteins. *Dev. Cell* **9**, 109–119 (2005).
19. Hayashi, K. *et al.* Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling. *Proc. Natl Acad. Sci. USA* **105**, 5632–5637 (2008).
20. Dhonukshe, P., Mathur, J., Hulskamp, M. & Gadella, T. W. Jr. Microtubule plus-ends reveal essential links between intracellular polarization and localized modulation of endocytosis during division-plane establishment in plant cells. *BMC Biol.* **3**, 11 (2005).
21. Karahara, I. *et al.* The preprophase band is a localized center of clathrin-mediated endocytosis in late prophase cells of the onion cotyledon epidermis. *Plant J.* **57**, 819–831 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Estelle, G. Wasteneys, K. Hayashi, C. Luschig, I. Hwang, M. Sasabe and M. Heisler for sharing published material. We thank F. Kindt and R. Leito for photography and image processing, and the Akhmanova laboratory for the western blotting facility. This work was supported by the Utrecht University's Starting Independent Investigator, the Netherlands Organisation for Scientific Research VIDI and European Research Council Starting Investigator grants to P.D., and the European Research Council Advanced Investigator grant, the Netherlands Organisation for Scientific Research Spinoza and Netherlands Proteomics Centre grants to B.S. K.K. was recruited from grants of P.D., and H.Z. was supported by grants of B.S.

Author Contributions P.D. conceptualized the project, designed the experiments and directed the work; B.S. contributed to the experimental design; K.K. and P.D. performed experiments and analysed data; H.Z. contributed to western analysis; P.D., K.K. and B.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D. (pankaj.dhonukshe@psb.ugent.be).

METHODS

Plant material and growth conditions. *Arabidopsis* mutants *clasp-1* (ref. 15), *pid wag1 wag2* (ref. 6), *pin2 pin3 pin7* (ref. 8), *pin2 pin4 pin7* (ref. 8), *tir1-1 afb2-1 afb3-1* (ref. 18), *scr²²*, *wei-tar²³*, *map65-1 map65-2* (ref. 24), *35S::MAP65-1,2 amiRNA³*, and transgenic marker lines *DR5::YFP-NLS²⁵*, *WER::XVE > iaaH³*, *PIN2::PIN2-GFP²⁶*, *PIN1::PIN1-GFP²⁷*, *PIN2::PIN1-GFP2* (ref. 9), *PID::PID-YFP²⁸*, *CO2::XVE > PID-YFP*, *G1090::XVE-PID⁹*, *35S::PLT2-GR²⁹* and *35S::CLC-GFP³⁰* have been described previously. *35S::PIN2-GFP* was created by assembling entry vectors containing the 35S promoter, *PIN2-GFP* and NOS terminator into the destination vector. Oestradiol-inducible *G1090::XVE > GFP-CLASP* was created by assembling entry vectors containing the *G1090::XVE* promoter, *GFP-CLASP* and NOS terminator into the destination vector. Transgenic marker lines in the *clasp-1* background were generated by crossing. Homozygous mutant plants were then selected from the F₂ segregating population based on the genotyping, the typical *clasp-1* phenotype, and the fluorescence of the marker. All seeds were surface-sterilized by gas sterilization protocol for 2 h, and plated onto Murashige–Skoog (MS) medium, supplemented with 1% sucrose, and 9 g l⁻¹ bacto-agar, pH 5.8. The plates were placed for the 24-h stratification period at 4 °C in the dark, then moved to a growth chamber (16-h light/8-h dark photoperiod at 23 °C) and kept vertical. All experiments were performed in at least three independent biological replicates.

PIN polarity. For PIN polarity quantification, the cells within the root apical meristem zone were considered. Shootward PIN2 percentages were quantified by counting the cells with shootward PIN2 to the total number of cortical cells in the entire root apical meristem.

Columella organization. For quantification of columella organization defects, cells showing abnormal shapes and more than four borders were counted as defective.

Gravity stimulation. For gravity stimulation, 6-day-old seedlings were rotated by 90° anticlockwise for 3 h. For 48 h gravistimulation, plates were rotated by 90° every 12 h (clockwise–anticlockwise). To measure the root tip bending, plates were photographed before and after the reorientation using a digital camera (Nikon D300S). The angle of root tip bending was assessed using Adobe Photoshop (version CS4.0).

Chemical treatments and fluorescent dyes. IAA (Ducheva), NAA (Ducheva), NPA (Ducheva), IAM (Ducheva), BFA (Sigma), CHX (Sigma), oestradiol (Sigma), oryzalin (Sigma), PEO-IAA (a gift from K. Hayashi), and DEX (Sigma) were used from dimethylsulphoxide-dissolved stock solutions at 10 μM IAA, 10 μM NAA, 10 μM NPA, 10 μM IAM, 50 μM BFA, 50 μM CHX, 10 μM oestradiol, 2 μM oryzalin, 20 μM PEO-IAA and 10 μM DEX working concentrations for indicated periods. FM4-64 (Molecular Probes) dissolved in water was applied at a 2 μM final concentration for 30 min. Propidium iodide (Sigma) dissolved in water was applied to roots for imminent analysis. For auxin treatment, 5-day-old wild-type and *clasp-1* seedlings were transferred onto MS media with 10 μM IAA, 10 μM NAA, 10 μM NPA or 10 μM IAM for 48 h. Wild-type and *clasp-1* seedlings expressing *PIN2::PIN2-GFP* and *PIN1::PIN1-GFP* were treated in liquid medium with 50 μM CHX for 30 min, followed by a 30-min treatment with 50 μM BFA and 50 μM CHX. After BFA and CHX treatments, seedlings were transferred to MS liquid medium for a further 1-h washout in the presence of 50 μM CHX. For *CO2::XVE > PID-YFP* and *G1090::XVE-PID* induction, seedlings were transferred to MS media with 10 μM oestradiol for 24 h. For columella organization and gravity analyses after PID induction, seedlings were germinated on 10 μM oestradiol-enriched media plates, and grown for up to 7 days. Oryzalin was applied to seedlings at 2 μM in MS liquid medium for 1 or 4 h. For PLT2-GR induction, seedlings were transferred to MS media enriched with 10 μM DEX for 16–24 h. For NPA-triggered cell division switch analysis, seedlings were transferred to MS media enriched with 10 μM NPA for 60 h. To observe *PID::PID-YFP* localization in response to PEO-IAA, its active and inactive forms were applied to seedlings at 20 μM in liquid MS media for 8 h. For immunolocalization analysis, PEO-IAA was applied to seedlings for 24 h. For columella organization, PIN2 relocation and root apical meristem size observations after CLASP induction in *clasp-1*, seedlings were transferred to MS plates enriched with 10 μM oestradiol for the indicated periods.

Starch granule staining. Six-day-old seedling roots were placed in Lugol's solution (Merck) for 2 min, then placed in chloral hydrate solution (80 g chloral hydrate, 20 ml glycerol, 20 ml water). Differential interface contrast (DIC) microscopy analysis was performed immediately.

EdU staining. Six-day-old seedlings of wild type, *clasp-1*, *map65-1 map65-2*, *tir1 afb2-1 afb3-1* and *CO::XVE > PID-YFP* were processed using Click-iT RNA Alexa Fluor 488 HCS assay, as described previously³¹. EdU-labelled root tips were immediately viewed by confocal laser scanning microscopy.

Immunolocalization. Whole-mount immunolocalization was performed on wild type, *clasp-1*, *pid wag1 wag2*, *scr*, *wei-tar*, as well as *clasp-1* mutants after BFA treatment and BFA washout. Furthermore, on wild type, *clasp-1*, *pid wag1 wag2*, *G1090::XVE-PID*, *35S::PLT2-GR*, *tir1-1 afb2-1 afb3-1*, *35S::MAP65-1,2 amiRNA*,

35S::PLT2-GR in *clasp-1*, *tir1-1 afb2-1 afb3-1*, *35S::MAP65-1,2 amiRNA*, *35S::PLT2-GR* and *35S::CLC-GFP* after PEO-IAA treatment. Seven-day-old seedlings were fixed in 4% paraformaldehyde in MTSB buffer (50 mM PIPES, 5 mM MgSO₄, 5 mM EGTA, pH 6.9) as described previously³². Before their analyses, PLT2-GR was induced for 16 h with 10 μM DEX. Rabbit-anti-PIN2 (1:800) primary antibody³³ or rabbit-anti-PIN1 (1:1,000) primary antibody (SCBT), in combination with Cy3-conjugated anti-rabbit (1:500) secondary antibody was used for PIN2 or PIN1 detection, respectively. Rabbit-anti-clathrin (1:400) primary antibody³⁴ in combination with either Cy3-conjugated anti-rabbit (1:500) secondary antibody (Sigma) or FITC-anti-rabbit (1:300) secondary antibody (Sigma) were used for clathrin heavy chain detection. Cells were then viewed by confocal laser scanning-microscopy.

Microscopy and image analysis. Confocal laser scanning-microscopy was performed using Leica TCS SPII confocal microscope. Propidium iodide, GFP, YFP, FM4-64, Cy3 and FITC signals were excited using the Argon laser and collected at 637–672, 498–523, 525–546, 585–663, 553–663 and 488–505 nm, respectively. All images were captured at ×63 magnification. Finally, images were processed using Adobe Photoshop (version CS4.0). Fluorescence signal intensity was measured using the Leica (Live) confocal software in a region of interest. The polar and lateral fluorescence intensity values for PIN, PID, clathrin and microtubules were obtained by manually drawing lines on the lateral and polar cell sides (yellow and light blue dotted lines in Supplementary Fig. 2b). The polar and lateral fluorescence intensity ratios for PIN, PID, clathrin and microtubules were obtained by averaging the polar to lateral fluorescence intensity ratios of individual cells. The plasma membrane fluorescence intensity values for PIN2, PID, clathrin and FM4-64 were obtained by manually drawing lines on all cell sides (yellow dotted lines in Supplementary Fig. 2a, c, d). Fluorescence intensity values for the internalized FM4-64 and PIN2 were obtained by manually drawing closed regions covering entire intracellular areas (light blue enclosed regions in Supplementary Fig. 2c, d). The percentage of internalized PIN2 (internalized PIN2/(internalized PIN2 + PIN2 at the plasma membrane) × 100) and internalized FM4-64 (internalized FM4-64/(internalized FM4-64 + FM4-64 at the plasma membrane) × 100) were obtained by averaging the internalized to total fluorescence intensity ratios of PIN2 and FM4-64 for individual cells. The percentage of recycled PIN2 (PIN2 at plasma membrane/(PIN2 at plasma membrane + PIN2 inside cell)) was obtained by averaging the recycled to total fluorescence intensity ratios of PIN2 for individual cells. Similar methodology was followed for quantifying PIN1 internalization and recycling. For DIC microscopy, seedlings were placed into chloral hydrate solution and were observed with a Zeiss Axioplan II microscope equipped with DIC optics and a CCD camera. All DIC images were taken at ×20 and ×40 magnifications.

Total RNA isolation and cDNA synthesis. Total RNA was extracted from six-day-old wild-type and *clasp-1* seedlings using the TRIZOL reagent (Sigma) following manufacturer's instructions. Three micrograms of total RNA was digested with RNase-free DNase I (Fermentas). The absence of contaminating genomic DNA after DNase I treatment was confirmed by PCR analysis using primer pairs (FLC1-intron-F, FLC1-intron-R) designed to amplify a 200-base-pair (bp) intron sequence of the control gene, *FLC1*. For cDNA synthesis, 1 μg of total RNA was primed with odT18 primer (Biolegio) and reverse transcribed using RevertAID M-MuLV reverse transcriptase (Qiagen). The cDNA concentration was assessed either by amplification of a 400–600-bp segment of *ACTIN 1* cDNA (ACT1-F, ACT1-R), or a 50-bp segment of *ACTIN 2* cDNA (ACT2-1154-F, ACT2-1204-R). Only cDNAs with equal concentrations were considered suitable for RT-PCR and qRT-PCR analyses. For all primer sequences see Supplementary Table 1.

RT-PCR analysis. A standard PCR protocol was applied to amplify all templates, followed by electrophoresis on 1.5% (v/w) agarose gel. For all primer sequences see Supplementary Table 1.

qRT-PCR analysis. qRT-PCR analysis was carried out in ABI PRISM 7900 HT sequence detection system (Applied Biosystems) using SYBR Green to monitor double-stranded DNA synthesis. Reactions of 20 μl total volume consisted of 2× Fast SYBR Green reagent (Applied Biosystems), 20 ng cDNA and 200 nM of each gene-specific primer. The following standard PCR protocol was applied to amplify all templates: 50 °C for 2 min; 95 °C for 10 min; 40 cycles of 95 °C for 15 s and 60 °C for 1 min, and SYBR Green fluorescence was measured continuously. Fluorescence data for melting curves were acquired after 40 cycles by heating the sample up to 95 °C for 15 s, followed by cooling down to 60 °C for 15 s, and heating the samples to 95 °C for 15 s. Data analysis was performed with the SDS 2.2.1 software (Applied Biosystems). To determine the threshold cycle value (C_t) for each PCR reaction, the threshold (ΔRn) was set within the logarithmic amplification phase. PCR efficiency (E) was estimated using LinReg software with data obtained from the exponential phase of each individual amplification plot and the equation: 1 + E = 10/slope. Relative expression levels of genes of interest are given in

comparison to *ACTIN 2* (ACT2-1154-F, ACT2-1204-R), and were calculated using the following formula: $\Delta\Delta C_t = \text{power}(1 + E; -\Delta C_t)$, in which $\Delta C_t = C_{t(\text{gene})} - C_{t(\text{ACT2})}$. For all primer sequences see Supplementary Table 1.

Protein extraction and western blot. Five-day-old seedling was frozen in liquid nitrogen and ground into fine powder. The extraction buffer (50 mM HEPES-KOH, pH 7.5, 100 mM KCl, 5 mM EDTA, 5 mM EGTA, 50 mM NaF, 1% (v/v) IGEPAL, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 1 mM phenylmethylsulfonyl-fluoride (PMSF), 1× proteinase inhibitor cocktail (Roche), 1× phosphSTOP(Roche)) was added to the powder to extract the protein. Protein extracts were spun twice for 15 min at 16,000g at 4 °C and soluble protein was quantified with Bio-Rad protein assay reagent. Twenty micrograms of protein was separated by 10% (w/v) SDS-PAGE and transferred to nitrocellulose ECL membrane (Fischer Scientific). Western blots were probed with anti-YFP antibody (Roche) or anti-actin C4 monoclonal antibody (Millipore).

Accession numbers. The *Arabidopsis* Genome Initiative locus identifiers for the genes mentioned in this manuscript are as follows: *CLASP* (At2g20190), *PIN1* (At1g73590), *PIN2* (At5g57090), *PID* (At2g34650), *WAG1* (At1g53700), *WAG2* (At3g14370), *MAP65-1* (At5g55230), *MAP65-2* (At4g26760), *TIR1* (At1g72930), *AFB2* (At3g26810), *AFB3* (At1g12820), *PLT2* (At1g51190), *CHC-1* (At3g08530), *CHC-2* (At3g11130), *CLC-2* (At2g40060), *SCR* (At3g54220), *WEI-TAR* (Atg1g-70560), *ACTIN 1* (At2g37620) and *ACTIN 2* (At3g18780).

22. Di Laurenzio, L. *et al.* The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* **86**, 423–433 (1996).
23. Stepanova, A. N. *et al.* TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* **133**, 177–191 (2008).
24. Sasabe, M., Kosetsu, K., Hidaka, M., Murase, A. & Machida, Y. *Arabidopsis thaliana* MAP65-1 and MAP65-2 function redundantly with MAP65-3/PLEIADE in cytokinesis downstream of MPK4. *Plant Signal. Behav.* **6**, 743–747 (2011).
25. Heisler, M. G. *et al.* Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the *Arabidopsis* inflorescence meristem. *Curr. Biol.* **15**, 1899–1911 (2005).
26. Xu, J. & Scheres, B. Dissection of *Arabidopsis* ADP-RIBOSYLATION FACTOR 1 function in epidermal cell polarity. *Plant Cell* **17**, 525–536 (2005).
27. Benková, E. *et al.* Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115**, 591–602 (2003).
28. Michniewicz, M. *et al.* Antagonistic regulation of PIN phosphorylation by PP2A and PINOID directs auxin flux. *Cell* **130**, 1044–1056 (2007).
29. Galinha, C. *et al.* PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* **449**, 1053–1057 (2007).
30. Konopka, C. A., Backues, S. K. & Bednarek, S. Y. Dynamics of *Arabidopsis* dynamin-related protein 1C and a clathrin light chain at the plasma membrane. *Plant Cell* **20**, 1363–1380 (2008).
31. Kotogány, E., Dudits, D., Horvath, G. V. & Ayaydin, F. A rapid and robust assay for detection of S-phase cell cycle progression in plant cells and tissues by using ethynyl deoxyuridine. *Plant Methods* **6**, 5 (2010).
32. Sauer, M., Paciorek, T., Benkova, E. & Friml, J. Immunocytochemical techniques for whole-mount *in situ* protein localization in plants. *Nature Protocols* **1**, 98–103 (2006).
33. Abas, L. *et al.* Intracellular trafficking and proteolysis of the *Arabidopsis* auxin-efflux facilitator PIN2 are involved in root gravitropism. *Nature Cell Biol.* **8**, 249–256 (2006).
34. Kim, Y. W. *et al.* *Arabidopsis* dynamin-like 2 that binds specifically to phosphatidylinositol 4-phosphate assembles into a high-molecular weight complex *in vivo* and *in vitro*. *Plant Physiol.* **127**, 1243–1255 (2001).

Conformational biosensors reveal GPCR signalling from endosomes

Roshanak Irannejad¹, Jin C. Tomshine¹, Jon R. Tomshine¹, Michael Chevalier², Jacob P. Mahoney³, Jan Steyaert^{4,5}, Søren G. F. Rasmussen⁶, Roger K. Sunahara³, Hana El-Samad², Bo Huang^{2,7} & Mark von Zastrow^{1,8}

A long-held tenet of molecular pharmacology is that canonical signal transduction mediated by G-protein-coupled receptor (GPCR) coupling to heterotrimeric G proteins is confined to the plasma membrane. Evidence supporting this traditional view is based on analytical methods that provide limited or no subcellular resolution¹. It has been subsequently proposed that signalling by internalized GPCRs is restricted to G-protein-independent mechanisms such as scaffolding by arrestins^{2,3}, or GPCR activation elicits a discrete form of persistent G protein signalling^{4–9}, or that internalized GPCRs can indeed contribute to the acute G-protein-mediated response¹⁰. Evidence supporting these various latter hypotheses is indirect or subject to alternative interpretation, and it remains unknown if endosome-localized GPCRs are even present in an active form. Here we describe the application of conformation-specific single-domain antibodies (nanobodies) to directly probe activation of the β_2 -adrenoceptor, a prototypical GPCR¹¹, and its cognate G protein, G_s (ref. 12), in living mammalian cells. We show that the adrenergic agonist isoprenaline promotes receptor and G protein activation in the plasma membrane as expected, but also in the early endosome membrane, and that internalized receptors contribute to the overall cellular cyclic AMP response within several minutes after agonist application. These findings provide direct support for the hypothesis that canonical GPCR signalling occurs from endosomes as well as the plasma membrane, and suggest a versatile strategy for probing dynamic conformational change *in vivo*.

Ligand binding to the extracellular surface of the β_2 -adrenoceptor (β_2 -AR) stabilizes an activating conformational change in the receptor that promotes guanine nucleotide dissociation from the cytoplasmic GTP-binding protein G_s ; this represents the critical biochemical event initiating classical GPCR signal transduction (Fig. 1a)¹³. Activated β_2 -ARs are substrates for phosphorylation and binding of β -arrestins, events which inhibit interaction with G proteins and promote endocytosis of receptors via clathrin-coated pits (CCPs)^{14,15}. Acute β_2 -AR G_s signalling is thus traditionally thought to be restricted to the plasma membrane^{14,16,17}. However, to our knowledge, this assumption has not been directly tested. To do so, we generated a biosensor of activated β_2 -AR based on a conformation-specific single-domain camelid antibody (Nb80) used in recent structural studies^{18,19}. We reasoned that this nanobody, which selectively binds the agonist-occupied β_2 -AR and is able to stabilize an activated receptor conformation when present *in vitro* at high concentration, might act as a sensor of receptor activation when expressed at relatively low concentration in intact cells (Fig. 1b). This proved to be the case; in cells maintained in the absence of agonist, Nb80 fused to enhanced green fluorescent protein (Nb80-GFP) localized to the cytoplasm and not with β_2 -ARs present in the plasma membrane (Fig. 1c, 0 min, top; Pearson's coefficient = 0.135). Line scan analysis verified the cytoplasmic distribution of Nb80-GFP before β_2 -AR activation (Fig. 1d, top) as expected because the

cytoplasmic concentration of Nb80-GFP achieved in our experiments (approximately 20 nM) was considerably lower than the equilibrium dissociation constant estimated *in vitro* for Nb80 binding to purified β_2 -ARs in the absence of agonist ($0.76 \pm 0.14 \mu\text{M}$; Supplementary Fig. 1a–d). After application of the adrenergic agonist isoprenaline (10 μM), Nb80-GFP was rapidly recruited to the plasma membrane and co-localized there with β_2 -ARs (Fig. 1c, middle; Pearson's coefficient = 0.625). Line scan analysis verified robust Nb80-GFP recruitment to the plasma membrane and concomitant depletion from the cytoplasm (Fig. 1d, middle), consistent with the much higher affinity of Nb80 for isoprenaline-activated β_2 -ARs ($2.9 \pm 0.5 \text{ nM}$; Supplementary Fig. 1d). Agonist-induced membrane recruitment of Nb80-GFP was specific because the D1 dopamine receptor (DRD1), which is also G_s -coupled but does not bind Nb80 *in vitro* (data not shown), failed to recruit Nb80-GFP to the plasma membrane in response to dopamine (10 μM) application (Supplementary Fig. 2). Furthermore, β_2 -AR-cyan fluorescent protein (CFP) and Nb80-yellow fluorescent protein (YFP) generated a pronounced fluorescence (Förster) resonance energy transfer (FRET) signal after isoprenaline application whereas DRD1-CFP did not (Supplementary Fig. 3a, b).

β_2 -AR internalization began 1 to 2 min after Nb80-GFP recruitment to the plasma membrane, indicated by the emergence of surface-labelled β_2 -AR in peripheral cytoplasmic vesicles. Nb80-GFP did not co-localize with β_2 -AR-containing endocytic vesicles upon first appearance (Fig. 1c, middle, arrow in merged image points to an example) but was recruited at later time points (Fig. 1c, bottom, Pearson's coefficient = 0.702; examples are indicated by arrowheads). Endosome recruitment of Nb80-GFP was evident by line scan analysis (Fig. 1d, bottom; line scans are from the representative individual examples with further quantification in legend) and localized to EEA1-marked early endosomes (Pearson's coefficient = 0.846; Supplementary Fig. 4) through which β_2 -ARs iteratively cycle in the presence of agonist²⁰. β_2 -AR-containing endosomes were initially devoid of Nb80-GFP and later acquired Nb80-GFP during their movement (Supplementary Videos 1 and 2). Interaction at endosomes was verified by β_2 -AR-CFP and Nb80-YFP normalized FRET (nFRET) (Supplementary Fig. 3c). These results suggest that β_2 -AR activation initiates a precisely choreographed series of events: Nb80-GFP is first recruited from the cytoplasm to the plasma membrane, then β_2 -ARs internalize devoid of Nb80-GFP, followed by a second phase of Nb80-GFP recruitment to the internalized β_2 -ARs.

Nb80-GFP recruitment to endosomes required β_2 -ARs because a phosphorylation-deficient mutant version of the β_2 -AR (β_2 -AR-3S, with three serine mutations) that couples to G_s but is impaired in agonist-induced endocytosis²¹ recruited Nb80-GFP to the plasma membrane but produced much less recruitment to endosomes (Fig. 1e, top). Nb80-GFP co-localized with β_2 -AR-3S after agonist-induced activation (Pearson's coefficient = 0.674) but this was largely restricted to the

¹Department of Psychiatry, University of California, San Francisco, California 94158, USA. ²Department of Biochemistry & Biophysics, University of California, San Francisco, California 94158, USA.

³Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁴Department of Molecular and Cellular Interactions, Vrije Universiteit Brussel, B-1050 Brussels, Belgium. ⁵Structural Biology Research Centre, VIB, B-1050 Brussels, Belgium. ⁶Department of Neuroscience and Pharmacology, The Panum Institute, University of Copenhagen, 2200 Copenhagen N, Denmark. ⁷Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, USA. ⁸Department of Cellular & Molecular Pharmacology, University of California, San Francisco, California 94158, USA.

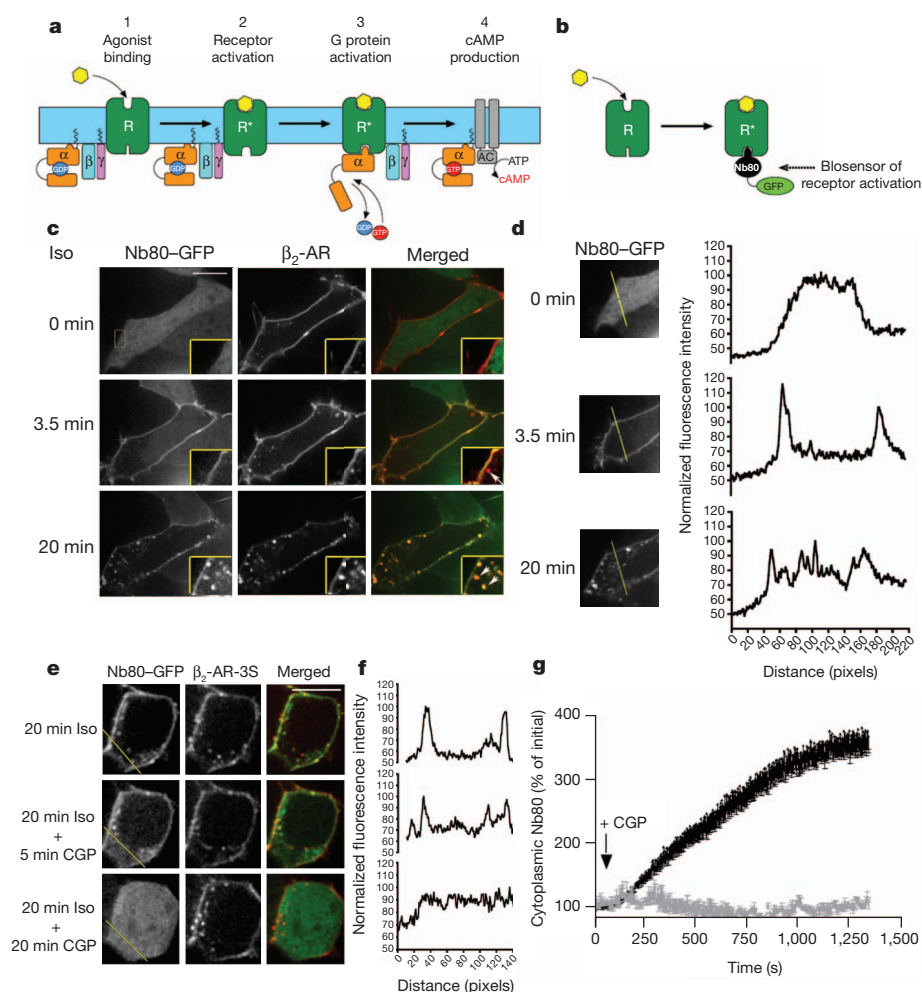


Figure 1 | Nb80-GFP detects activated β_2 -ARs in the plasma membrane and endosomes. **a**, The main events in β_2 -AR cAMP signalling include agonist binding (step 1), conformational activation of the receptor (step 2) that is coupled to conformational activation of G_s (step 3) that produces guanine nucleotide exchange on G_s and subsequent activation of adenylyl cyclase (AC) (step 4). **b**, Scheme for detecting conformational activation of β_2 -AR with Nb80-GFP. **c**, Representative Nb80-GFP (green) and β_2 -AR (red) localization at the indicated time (left) after 10 μ M isoprenaline addition (>30 Nb80-GFP positive endosomes per cell observed at 20 min; $n = 29$ cells, 10 experiments). **d**, Representative individual Nb80-GFP line scans (shown at the same magnification as panel **c**). **e**, Representative Nb80-GFP (green) and β_2 -AR-3S (red) localization after 20 min of isoprenaline treatment (top) followed by reversal with 50 μ M CGP-12177 for the indicated times (6.4 Nb80-GFP positive endosomes per cell; $n = 40$ cells, 3 experiments). **f**, Representative individual Nb80-GFP line scans. **g**, Recovery of cytoplasmic Nb80-GFP fluorescence (black) or bleaching control of the plasma membrane β_2 -AR-3S (grey) (mean \pm s.e.m., $n = 5$ experiments). Scale bars, 10 μ m.

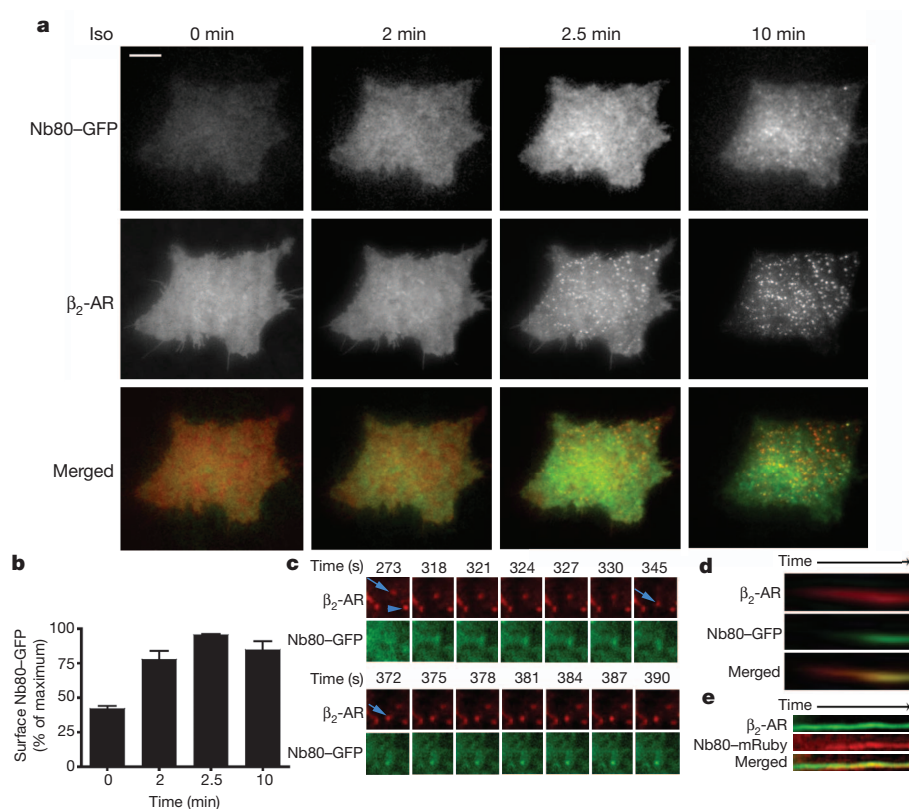


Figure 2 | Nb80-GFP accumulates on β_2 -AR-containing endosomes after their formation. **a**, β_2 -AR (red) and Nb80-GFP (green) at the indicated times after isoprenaline addition. Scale bar, 10 μ m. **b**, Average Nb80-GFP fluorescence measured in the TIRF illumination field at the indicated times (mean \pm s.e.m., $n = 7$ cells). **c**, TIRF image series showing β_2 -AR (red) and Nb80-GFP (green) in sequential frames. **d**, Kymograph of an individual β_2 -AR-containing endosome (red, Alexa555) showing Nb80-GFP (green) acquisition over 4 min. **e**, Kymograph of β_2 -AR (green, Alexa488) and Nb80-mRuby (red) over 6 min.

plasma membrane (line scan analysis in Fig. 1f is representative; further quantification in the figure legend). Nb80–GFP membrane recruitment was also reversible because the biosensor returned to a cytoplasmic distribution after addition of the competitive antagonist CGP-12177 (Fig. 1e, f, middle and bottom rows; Pearson's coefficient = 0.106), verified by recovery of Nb80–GFP fluorescence intensity in the cytoplasm (Fig. 1g, see also Supplementary Video 3). Thus, Nb80–GFP detected activated β_2 -ARs both in the plasma membrane and endosomes after acute agonist application.

Discrete Nb80–GFP recruitment phases were clearly resolved by total internal reflection fluorescence (TIRF) microscopy that selectively detects events occurring in the plasma membrane and extending approximately 100 nm into the peripheral cytoplasm²². First, within 2 min after agonist application, Nb80–GFP was progressively recruited to the plasma membrane in a diffuse distribution (Fig. 2a, compare first and second columns from left; quantification in Fig. 2b). Nb80–GFP did not co-localize with β_2 -ARs when they clustered in relatively static punctae characteristic of receptor-containing CCPs²³ (Fig. 2a, third column). Second, over a period of several additional minutes, Nb80–GFP was recruited to a discrete population of highly mobile punctae (Fig. 2a, fourth column) representing peripheral β_2 -AR-containing endocytic vesicles²³.

Sequential TIRF imaging emphasized the distinction between relatively static β_2 -AR puncta not co-localizing with Nb80–GFP (Fig. 2c, arrowhead indicates an example) and mobile endosomes that did (Fig. 2c, arrow indicates a representative example; many are visible in Supplementary Video 4). Later recruitment of Nb80–GFP occurred rather suddenly, typically within approximately 5 s (Supplementary Video 5), as also evident in kymographs of individual endosome trajectories (Fig. 2d). Ruling out potential artefacts of wavelength-dependent differences in TIRF microscopy illumination depth, later recruitment of the biosensor to endosomes was similarly observed when the excitation wavelengths used to detect receptor and biosensor were reversed (Fig. 2e).

Nb80–GFP did not detectably concentrate in CCPs labelled with either of two independent markers, the adaptor protein β -arrestin-2 (Fig. 3a and Supplementary Video 6) or the coat protein component clathrin light chain (Fig. 3b and Supplementary Video 7). Representative examples are shown and this was verified across multiple cells and experiments (Pearson's coefficient = 0.365 and 0.319, respectively, with numbers of replicates specified in the legend). In contrast, and as expected based on previous work, extensive co-localization of β_2 -AR with β -arrestin was observed under the same conditions (Pearson's coefficient = 0.677). Separation of Nb80–GFP localization from that of either β -arrestin or clathrin was clear in line scans (bottom panels of Fig. 3a, b were derived from the images shown and are representative, further quantification is in the legend).

A simple interpretation is that Nb80–GFP associates with activated β_2 -ARs in the plasma membrane but then dissociates before receptors cluster in CCPs and internalize. This was surprising because β_2 -AR clustering in CCPs occurs much more rapidly²³ than the overall reversal rate of Nb80–GFP recruitment observed after agonist washout in intact cells (Fig. 1g), or the kinetics of Nb80 dissociation from purified β_2 -ARs *in vitro* (Supplementary Fig. 1b, e). One possibility is that Nb80–GFP dissociation is accelerated during the clustering process, by mechanisms such as receptor phosphorylation or steric exclusion mediated by β -arrestins or other CCP-associated components. An alternative possibility is that the fraction of β_2 -ARs that have bound Nb80–GFP in the plasma membrane by the time of the clustering reaction are unable to enter CCPs, and only those β_2 -ARs not initially bound to Nb80–GFP in the plasma membrane are able to cluster in CCPs and subsequently internalize. In either case, the data clearly indicate that Nb80–GFP associates with β_2 -ARs after endocytosis, and after uncoating of the endocytic vesicle has occurred (Fig. 3c). Accordingly, Nb80–GFP recruitment to β_2 -AR-containing endosomes cannot represent an artefact of persistent nanobody binding from the plasma membrane; instead, this observation reveals that β_2 -ARs present in early endosomes are in an activated conformation.

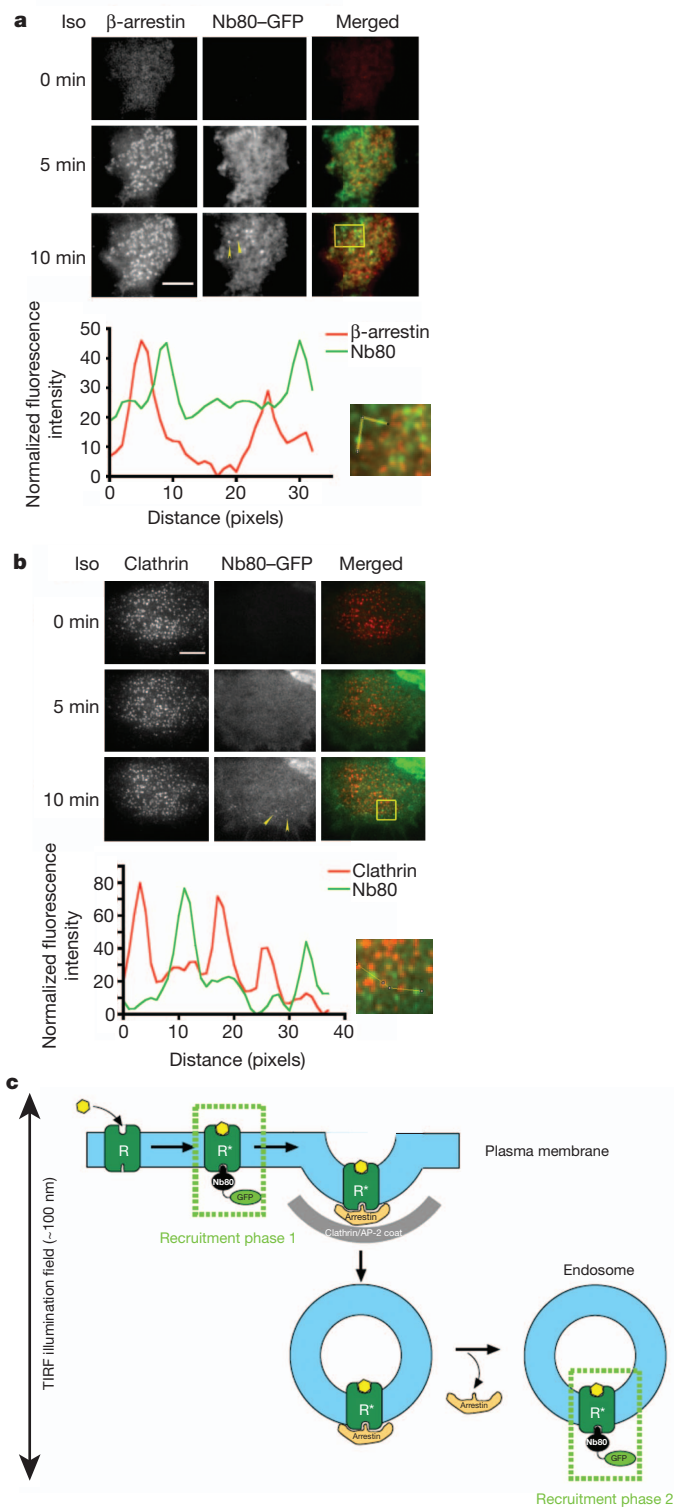


Figure 3 | Nb80–GFP does not accumulate in clathrin-coated pits or vesicles. **a**, Representative TIRF microscopy frames showing Nb80–GFP (green) and β -arrestin-2-mCherry (red) before and after agonist addition. Fluorescence intensity profiles are shown below for the indicated region and path; representative of $n = 39$ cells, 5 experiments and 4,849 punctae. **b**, Equivalent analysis comparing Nb80–GFP (green) to clathrin light chain-dsRed (red); representative of $n = 26$ cells, 3 experiments and 3,965 punctae. Arrowheads indicate examples of Nb80–GFP labelled endosomes. Scale bars, 5 μ m. **c**, Model for two phases of Nb80–GFP recruitment by the activated β_2 -AR, first at the plasma membrane and then at endosomes.

Because endosomes contain activated β_2 -ARs, we next asked if receptors engage their cognate G protein from this compartment. Heterotrimeric G proteins and adenylyl cyclase can be observed in endosomes as well as at the plasma membrane, supporting the concept of endosome-based G protein signalling^{4,6,10,24}. To directly investigate the subcellular location of G protein activation, we developed a distinct biosensor based on another nanobody, Nb37, which specifically recognizes the guanine-nucleotide-free form of $G\alpha_s$ representing the catalytic intermediate of G protein activation (Fig. 4a)²⁵. We hoped that, because Nb37 binds a surface of the alpha-helical domain that is accessible only in the nucleotide-free form, we would be able to detect production of this critical but fleeting activation intermediate in living cells. This was indeed the case because Nb37-GFP localized in the cytoplasm of untreated cells and was rapidly recruited from the cytoplasm to the plasma membrane in response to isoprenaline application (Fig. 4b, yellow arrowhead, Pearson coefficient = 0.627). Membrane recruitment of Nb37-GFP was agonist-dependent even in cells overexpressing $G\alpha_s$ and, notably, the cytoplasmic concentration of Nb37-GFP achieved in our experiments (also approximately 20 nM) was substantially lower than that producing detectable inhibition of G protein activation *in vitro* (Supplementary Fig. 5). Together, these observations suggest that Nb37-GFP indeed functions as a specific biosensor of G_s activation under the experimental conditions used.

Nb37-GFP was recruited not only to the plasma membrane (Fig. 4b, yellow arrowhead) but also to β_2 -AR-containing endosomes. Notably, endosome recruitment of Nb37-GFP occurred after the appearance of β_2 -ARs in the endosome. Such later recruitment was evident in dual label confocal image series (Fig. 4b shows an example: white arrowhead indicates recently internalized β_2 -AR, white arrows indicate Nb37-GFP recruitment, Pearson coefficient = 0.710; see also Supplementary Video

8). Nb37-GFP recruitment was specifically dependent on receptor activation because a mutant β_2 -AR that is defective in G-protein coupling (β_2 -AR-Cys 341 Gly)²⁶ did not produce detectable recruitment (Supplementary Fig. 6), whereas the distinct G_s -coupled DRD1 recruited Nb37-GFP to both the plasma membrane and endosomes (data not shown). Nb37-GFP localized uniformly on β_2 -AR-containing endosomes in cells overexpressing $G\alpha_s$ (Fig. 4b and Supplementary Video 8) but was recruited non-uniformly in cells expressing endogenous levels (Fig. 4c, arrowhead indicates an example). Moreover, Nb37-GFP recruitment to endosomes had a scintillating appearance, fluctuating rapidly when viewed in live image series (Supplementary Video 9), suggesting that G protein activation detected by Nb37-GFP occurs dynamically from limited regions of the endosome membrane.

We then asked if active β_2 -AR G_s coupling from endosomes contributes to the cellular response, focusing on cAMP as a classical second messenger carrying the downstream signal, using a previously described real-time assay of cAMP accumulation in living cells and normalizing to the signal produced by receptor-independent activation of adenylyl cyclase with forskolin (5 μ M)²⁷. Isoprenaline produced a rapid increase in cytoplasmic cAMP accumulation, reaching a maximum within approximately 10 min (Fig. 4d, blue line). Dyngo-4a, a dynamin inhibitor that blocks β_2 -AR endocytosis by inhibiting CCP function^{28,29}, did not detectably affect the earliest portion of the forskolin-normalized cAMP accumulation curve but significantly reduced its later rise (Fig. 4d, green line, grey dots indicate *P* values), consistent with the two phases of endosome-localized activation detected by the nanobody-derived biosensors. A similar inhibition was observed in cells expressing a mutant β_2 -AR with a carboxy-terminal alanine residue added that prevents efficient recycling

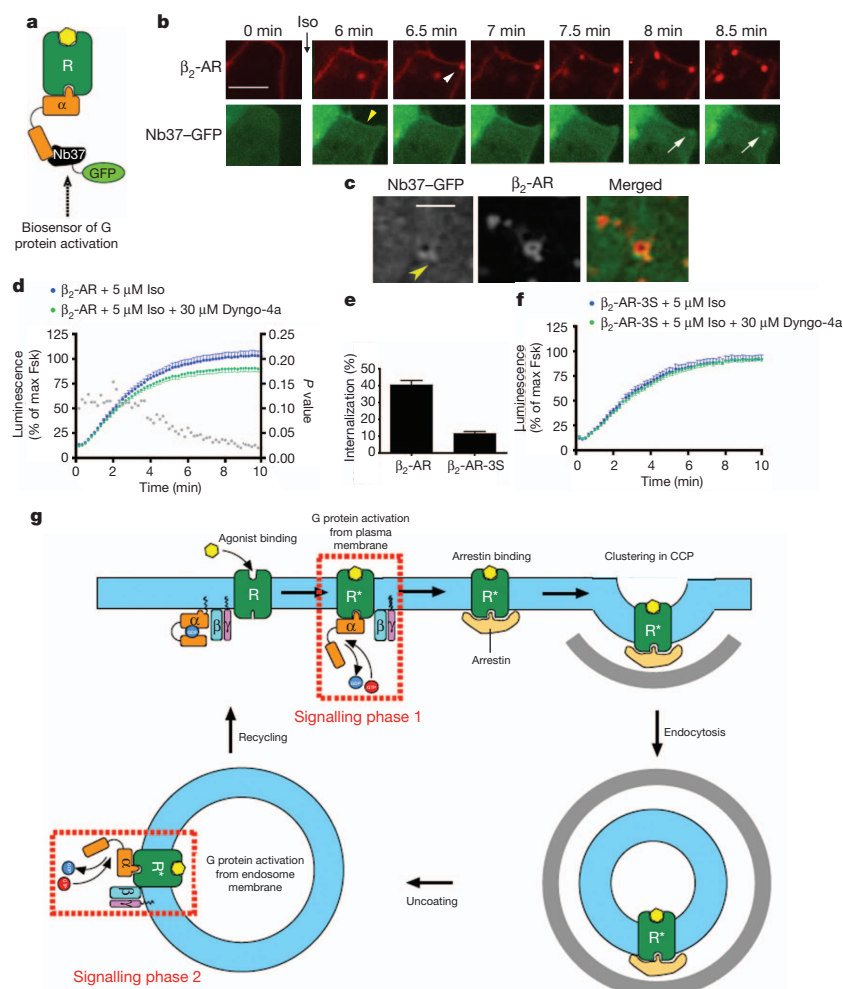


Figure 4 | Internalized β_2 -ARs contribute to the acute cAMP response. **a**, Scheme for detecting conformational activation of G_s with Nb37-GFP. **b**, Confocal image frames showing Nb37-GFP (green) and β_2 -AR (red) at the indicated time points after isoprenaline addition (representative of $n = 14$ cells; estimated Nb37-GFP recruitment delay ranged from 0.7 to 2.65 min). Scale bar, 5 μ m. Yellow arrowhead indicates Nb37-GFP recruitment to the plasma membrane. White arrowhead indicates an endosome containing recently internalized β_2 -AR and not associated with Nb37-GFP, white arrows indicate Nb37-GFP recruitment. **c**, Representative confocal frames showing a discrete endosomal structure labelled with Nb37-GFP (green) and β_2 -AR (red) at higher magnification. Scale bar, 2 μ m. Arrowhead indicates non-uniform localization of Nb37-GFP to the endosome. **d**, Forskolin-normalized β_2 -AR-mediated cAMP response in the absence or presence of 30 μ M Dyngo-4a (mean \pm s.e.m., $n = 10$ experiments, *P* values in grey). **e**, Isoprenaline (20 min)-induced β_2 -AR and β_2 -AR-3S internalization measured by flow cytometry ($n = 4$ experiments). **f**, Forskolin-normalized β_2 -AR-3S-mediated cAMP response in the absence or presence of 30 μ M Dyngo-4a ($n = 8$ experiments; $P = 0.1192$). **g**, Model for two phases of β_2 -AR G_s activation, first at the plasma membrane and then at endosomes, separated by the endocytic event.

(β_2 -AR-Ala)³⁰ (Supplementary Fig. 7), suggesting that the Dyngo-sensitive component of the cAMP response did not represent a secondary consequence of resensitization by receptor recycling¹⁷. Dyngo-4a did not produce the same effect on cAMP accumulation elicited by the internalization-defective β_2 -AR-3S mutant receptor²¹ (Fig. 4e, f), further supporting the conclusion that the second signalling phase indeed requires receptor localization in endosomes.

The present findings revise a long-held tenet of molecular pharmacology, that acute signal transduction mediated by the canonical β_2 -AR G_s signal transduction mechanism is plasma membrane delimited. The results provide direct evidence supporting the hypothesis that β_2 -AR endocytosis contributes to a second phase of the acute cellular cAMP response, which represents a significant component of the overall biochemical signal developed within several minutes after the initial agonist application. Thus, although it remains clear that β_2 -ARs can elicit G_s -mediated signal transduction from the plasma membrane, the present data reveal a discrete component of the acute signalling response that is initiated from endosomes (Fig. 4g). It remains unknown if β_2 -ARs are continuously bound by agonist in endosomes, as depicted in the figure for simplicity, but conformational activation of G_s in endosomes is both receptor- and agonist-dependent. Unambiguous detection of endosome-based activation of acute G-protein-linked signalling is presently limited to the β_2 -AR G_s system for which the critical nanobodies are available. However, endosome-based contribution to the acute signalling response is probably widespread in the GPCR superfamily because the β_2 -AR belongs to the largest group (family A) of GPCRs and is often considered a prototype. We also suggest, more generally, that nanobody-based biosensors represent a versatile strategy for probing other types of dynamic conformational change with high spatiotemporal resolution in living cells.

METHODS SUMMARY

Experiments were carried out using human embryonic kidney HEK293 cells (ATCC) expressing the indicated receptor constructs and nanobodies fused to enhanced GFP. Optical imaging was carried out at 37 °C in DMEM not containing phenol red and supplemented with 30 mM HEPES. Live cell cAMP accumulation was assessed at 37 °C using pGloSensor-20F (Promega). Flow cytometry was carried out using Alexa647 (Invitrogen)-conjugated M1 anti-Flag monoclonal antibody (Sigma) and a FACSCalibur instrument (Becton Dickinson).

Full Methods and any associated references are available in the online version of the paper.

Received 13 August 2012; accepted 8 February 2013.

Published online 20 March 2013.

1. Lohse, M. J., Nuber, S. & Hoffmann, C. Fluorescence/bioluminescence resonance energy transfer techniques to study G-protein-coupled receptor activation and signaling. *Pharmacol. Rev.* **64**, 299–336 (2012).
2. Shenoy, S. K. & Lefkowitz, R. J. Seven-transmembrane receptor signaling through β -arrestin. *Sci. STKE* **2005**, cm10 (2005).
3. Murphy, J. E., Padilla, B. E., Hasdemir, B., Cottrell, G. S. & Bunnett, N. W. Endosomes: a legitimate platform for the signaling train. *Proc. Natl Acad. Sci. USA* **106**, 17615–17622 (2009).
4. Ferrandon, S. et al. Sustained cyclic AMP production by parathyroid hormone receptor endocytosis. *Nature Chem. Biol.* **5**, 734–742 (2009).
5. Feinstein, T. N. et al. Retromer terminates the generation of cAMP by internalized PTH receptors. *Nature Chem. Biol.* **7**, 278–284 (2011).
6. Calebiro, D. et al. Persistent cAMP-signals triggered by internalized G-protein-coupled receptors. *PLoS Biol.* **7**, e1000172 (2009).
7. Werthmann, R. C., Volpe, S., Lohse, M. J. & Calebiro, D. Persistent cAMP signaling by internalized TSH receptors occurs in thyroid but not in HEK293 cells. *FASEB J.* **26**, 2043–2048 (2012).
8. Mullershausen, F. et al. Persistent signaling induced by FTY720-phosphate is mediated by internalized S1P1 receptors. *Nature Chem. Biol.* **5**, 428–434 (2009).
9. Calebiro, D., Nikolaev, V. O. & Lohse, M. J. Imaging of persistent cAMP signaling by internalized G protein-coupled receptors. *J. Mol. Endocrinol.* **45**, 1–8 (2010).
10. Kotowski, S. J., Hopf, F. W., Seif, T., Bonci, A. & von Zastrow, M. Endocytosis promotes rapid dopaminergic signaling. *Neuron* **71**, 278–290 (2011).
11. Lefkowitz, R. J. Seven transmembrane receptors: something old, something new. *Acta Physiol. (Oxf.)* **190**, 9–19 (2007).

12. Gilman, A. G. Transmembrane signaling, G proteins, and adenylyl cyclase. *Harvey Lect.* **85**, 153–172 (1989).
13. Rasmussen, S. G. et al. Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
14. Moore, C. A., Milano, S. K. & Benovic, J. L. Regulation of receptor trafficking by GRKs and arrestins. *Annu. Rev. Physiol.* **69**, 451–482 (2007).
15. Sorkin, A. & von Zastrow, M. Endocytosis and signalling: intertwining molecular networks. *Nature Rev. Mol. Cell Biol.* **10**, 609–622 (2009).
16. Luttrell, L. M. & Lefkowitz, R. J. The role of β -arrestins in the termination and transduction of G-protein-coupled receptor signals. *J. Cell Sci.* **115**, 455–465 (2002).
17. Gainetdinov, R. R., Premont, R. T., Bohn, L. M., Lefkowitz, R. J. & Caron, M. G. Desensitization of G protein-coupled receptors and neuronal functions. *Annu. Rev. Neurosci.* **27**, 107–144 (2004).
18. Rasmussen, S. G. et al. Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**, 175–180 (2011).
19. Steyaert, J. & Kobilka, B. K. Nanobody stabilization of G protein-coupled receptor conformational states. *Curr. Opin. Struct. Biol.* **21**, 567–572 (2011).
20. Lauffer, B. E. et al. SNX27 mediates PDZ-directed sorting from endosomes to the plasma membrane. *J. Cell Biol.* **190**, 565–574 (2010).
21. Hausdorff, W. P. et al. A small region of the β -adrenergic receptor is selectively involved in its rapid regulation. *Proc. Natl Acad. Sci. USA* **88**, 2979–2983 (1991).
22. Steyer, J. A. & Almers, W. A real-time view of life within 100 nm of the plasma membrane. *Nature Rev. Mol. Cell Biol.* **2**, 268–275 (2001).
23. Puthenveedu, M. A. & von Zastrow, M. Cargo regulates clathrin-coated pit dynamics. *Cell* **127**, 113–124 (2006).
24. Slessareva, J. E., Routt, S. M., Temple, B., Bankaitis, V. A. & Dohlman, H. G. Activation of the phosphatidylinositol 3-kinase Vps34 by a G protein α subunit at the endosome. *Cell* **126**, 191–203 (2006).
25. Westfield, G. H. et al. Structural flexibility of the $G_{s\alpha}$ helical domain in the β_2 -adrenoceptor G_s complex. *Proc. Natl Acad. Sci. USA* **108**, 16086–16091 (2011).
26. Campbell, P. T. et al. Mutations of the human β_2 -adrenergic receptor that impair coupling to G_s interfere with receptor down-regulation but not sequestration. *Mol. Pharmacol.* **39**, 192–198 (1991).
27. Violin, J. D. et al. β_2 -adrenergic receptor signaling and desensitization elucidated by quantitative modeling of real time cAMP dynamics. *J. Biol. Chem.* **283**, 2949–2961 (2008).
28. Harper, C. B. et al. Dynamin inhibition blocks botulinum neurotoxin type A endocytosis in neurons and delays botulism. *J. Biol. Chem.* **286**, 35966–35976 (2011).
29. Howes, M. T. et al. Clathrin-independent carriers form a high capacity endocytic sorting system at the leading edge of migrating cells. *J. Cell Biol.* **190**, 675–691 (2010).
30. Cao, T. T., Deacon, H. W., Reczek, D., Bretscher, A. & von Zastrow, M. A kinase-regulated PDZ-domain interaction controls endocytic sorting of the β_2 -adrenergic receptor. *Nature* **401**, 286–290 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Kobilka, P. Robinson, A. Kruse, E. Pardon, P. Temkin, M. Puthenveedu, A. Henry, A. Marley and K. Thorn for assistance, advice and discussion. These studies were supported by the National Institute on Drug Abuse of the US National Institutes of Health (DA010711 and DA012864 to M.v.Z. and F32 DA029993 to J.C.T.). R.I. is supported by the American Heart Association. R.K.S. and J.P.M. are supported by the National Institute of General Medical Sciences (GM083118 to R.K.S. and T32 GM007767 to J.P.M.). S.G.F.R. is supported by the Lundbeck Foundation. J.S. is supported by FWO-Vlaanderen grants (FWO551 and FWO646) and Innoviris-Brussels (BRGE02132). B.H. is supported by a Packard Fellowship for Science and Engineering.

Author Contributions R.I. constructed and validated the nanobody biosensors, carried out most of the cell biological experiments and analysis, contributed to overall experimental strategy and took a lead role in writing the manuscript. J.C.T. carried out early experiments identifying endocytic inhibitor effects on cellular cAMP signalling, and contributed to initial project planning. J.R.T. built the luminometer system, developed software for analysis of luminometry data, and contributed to early experiments on cellular cAMP signalling. M.C. contributed to experimental design and data analysis, and modelled effects of endocytic inhibitors on the cellular cAMP response. J.P.M. contributed to the production of receptor-containing rHDL particles and carried out *in vitro* studies of Nb80 binding and dissociation. J.S. developed the nanobody reagents used as the basis for the biosensors described in this study and advised on biosensor design and expression. S.G.F.R. contributed to developing and screening the initial nanobody reagents, and carried out *in vitro* studies of Nb80 binding and dissociation in rHDL particles reconstituted with bimane-labeled receptors. R.K.S. contributed to overall experimental interpretation, supervised J.P.M. in carrying out *in vitro* studies of Nb80 binding to receptors, and performed *in vitro* experiments evaluating Nb37 effects on G protein activation. H.E.-S. contributed to experimental design and data interpretation, and supervised efforts to model endocytic effects on the cellular cAMP response. B.H. contributed to overall experimental design and interpretation, implementation of biosensors and advised on image analysis. M.v.Z. was responsible for overall project strategy, carried out some of the imaging experiments, and drafted the manuscript together with R.I.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.v.Z. (Mark.VonZastrow@ucsf.edu).

METHODS

Cell Culture, cDNA constructs and transfection. HEK293 cells were grown in DMEM supplemented with 10% FBS (UCSF Cell Culture Facility) without antibiotics. Stably transfected HEK293 cell clones expressing Flag-tagged β_2 -AR-3S were created using previously described Flag-tagged β_2 -AR³⁰. A plasmid encoding a cyclic permuted luciferase reporter construct, based on a mutated RII β cAMP-binding domain from PKA (pGloSensor-20F, Promega). Nb80-eGFP and Nb37-eGFP were created by amplifying Nb80 and Nb37 nanobody complementary DNAs using 5'-CTTGAAAAGCTTGCCGCCACCATGGGACAGGTGCAGCTGCA-3'; 5'-TTCAAGGGATCCATGTGATGGTGTGATGGTGGTGTGAGGAGACGGT-3' and 5'-CTTGAAAAGCTTGCCGCCACCATGGGACAGGTGCAGCTGCA-3'; 5'-TTCAAGGGATCCATGTGATGGGCTTCAGGTTCTGTGATGGTGTGATGGT-3' primers, respectively, and cloning into the pEGFP-N1 vector using HindIII and BamHI. β -arrestin-2-GFP, clathrin-DsRed, EEA1-DsRed and G_{α_s} -HA were gifts from M. Caron, W. Almers, K. Mostov and P. Wedegaertner, respectively. β -arrestin-2-mCherry was generated by subcloning β -arrestin-2 to pmCherry (Clontech) and β_2 -AR-CFP was generated from the Flag-tagged β_2 -AR construct. Transfections were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Flag-tagged human β_2 -AR and β_2 -AR-3S (Ser 355 Gly, Ser 356 Gly and Ser 364 Gly were mutated simultaneously) constructs were labelled with Alexa555- or Alexa488-conjugated M1 anti-Flag monoclonal antibody (Sigma) as described previously³¹.

Live-cell confocal imaging. Live cell imaging was carried out using Yokagawa CSU22 spinning disk confocal microscope with a $\times 100$, 1.4 numerical aperture, oil objective and a CO₂ and 37 °C temperature-controlled incubator. A 488 nm argon laser and a 568 nm argon/krypton laser (Melles Griot) were used as light sources for imaging GFP and Flag signals, respectively. Cells expressing both Flag-tagged receptor and the indicated nanobody-GFP were plated onto glass coverslips. Receptors were surface labelled by addition of M1 anti-Flag antibody (1:1,000, Sigma) conjugated to Alexa 555 (A10470, Invitrogen) to the media for 30 min, as described previously³². Indicated agonist (isoprenaline, Sigma) or antagonist (CGP-12177, Tocris) were added and cells were imaged every 3 s for 20 min in DMEM without phenol red supplemented with 30 mM HEPES, pH 7.4 (UCSF Cell Culture Facility). Time-lapse images were acquired with a Cascade II EM charge-coupled-device (CCD) camera (Photometrics) driven by Micro-Manager 1.4 (<http://www.micro-manager.org>).

Live cell TIRF microscopy. TIRF imaging was carried out as described previously³³. Briefly, HEK293 cells co-expressing either Nb80-eGFP and β -arrestin-2-mCherry or clathrin light chain-DsRed, were imaged in DMEM without phenol red supplemented with 30 mM HEPES, pH 7.4 (UCSF Cell Culture Facility). Imaging was carried out using a Nikon TE-2000E inverted microscope with a $\times 100$, 1.49 numerical aperture TIRF objective, equipped for through-the-objective TIRF illumination, a 37 °C temperature-controlled stage (Bioscience Tools) and an objective warmer (Biopetechs). A 488 nm argon laser (Melles Griot) and a 543 nm helium-neon laser (Spectra Physics) were used as light sources. Time-lapse sequences were acquired with a C9100-12 camera (Hamamatsu Photonics) driven by iQ software (Andor). Cells were imaged every 3 s for 20 min.

Image analysis and statistical analysis. Images were saved as 16-bit TIFF files. Quantitative image analysis was carried out on unprocessed images using ImageJ software (<http://rsb.info.nih.gov/ij/>). Co-localization analysis was estimated by calculating the Pearson's correlation coefficient between the indicated image channels using the co-localization plug-in for ImageJ. Analysis of Nb80-GFP intensity profile along the straight line and Nb80-GFP/ β -arrestin or Nb80-GFP/clathrin along the segmented line were carried out using the ImageJ plot profile function. For estimating changes in Nb80-GFP surface fluorescence over time in TIRF images, individual cells were selected manually and fluorescence values measured over the entire stack. A blank area of the image lacking cells was used to estimate background fluorescence. Average fluorescence intensity was measured in each frame, background-subtracted and normalized to the maximum value. *P* values are from one-tailed unpaired Student's *t*-tests. For visual presentation (but not quantitative analysis), image series were processed using Kalman stack filter in ImageJ.

Luminescence-based rapid cAMP assay. HEK293 cells were transfected with a plasmid encoding a cyclic-permuted luciferase reporter construct, based on a mutated RII β cAMP-binding domain from PKA (pGloSensor-20F, Promega), which produces rapid and reversible cAMP-dependent activation of luciferase activity in intact cells. Cells were plated in 24-well dishes containing approximately 200,000 cells per well in 500 μ l DMEM without phenol red and no serum and equilibrated to 37 °C in a light-proof cabinet. An image of the plate was focused on a 512 \times 512 pixel electron multiplying CCD sensor (Hamamatsu C9100-13), cells were equilibrated for 1 h in the presence of 250 μ g ml⁻¹ luciferin (Biogold), and sequential luminescence images were collected every 10 s to obtain basal luminescence values. The camera shutter was closed, the cabinet opened and

the indicated concentration of isoprenaline was bath applied, with gentle manual rocking before replacing in the dark cabinet and resuming luminescence image acquisition. In endocytic manipulation experiments, cells were pre-incubated with 30 μ M Dyngo-4a (abcam Biochemicals) for 15 min. Every 10 s, sequential images were acquired using Micro-Manager (<http://www.micro-manager.org>) and integrated luminescence intensity detected from each well was calculated after background subtraction and correction for vignetting using scripts written in MATLAB (MathWorks). In each multiwell plate, and for each experimental condition, a reference value of luminescence was measured in the presence of 5 μ M forskolin, a manipulation that stimulates a moderate amount of receptor-independent activation of adenylyl cyclase. The average luminescence value—measured across duplicate wells—was normalized to the maximum luminescence value measured in the presence of 5 μ M forskolin.

FRET imaging. FRET imaging was carried out as described previously¹⁰. Briefly, HEK293 cells co-expressing β_2 -AR-CFP or DRD1-CFP and Nb80-YFP were imaged in wide field at 37 °C using a shuttered mercury arc lamp and standard CFP excitation (ET430/24 \times) and YFP emission (ET500/20 \times) band pass filters (Chroma). YFP emission was collected using a 535/30 m filter, and CFP emission was collected through a 470/24 m filter. Corrected FRET ratios were obtained using the following equation: $\text{NFRET} = [(I_{\text{FRET}} - \text{BG}_{\text{FRET}}) - (I_{\text{CFP}} - \text{BG}_{\text{CFP}}) \times \text{BT}_{\text{DONOR}} - (I_{\text{YFP}} - \text{BG}_{\text{YFP}}) \times \text{DE}_{\text{ACCEPTOR}}] / I_{\text{CFP}} \times \text{BT}_{\text{DONOR}}$, donor bleed through; $\text{DE}_{\text{ACCEPTOR}}$, direct excitation of the acceptor; BG_X , background fluorescence; and I_X , integrated fluorescence intensity measured in a given channel.

Flow cytometric assay of receptor endocytosis. Surface fluorescence of Flag- β_2 -AR or Flag- β_2 -AR-3S expressing HEK293 cells was used to measure receptor endocytosis. Cells were incubated with 10 μ M isoprenaline for 20 min at 37 °C to drive receptor internalization to steady state and were subsequently rinsed 3 times with ice-cold PBS, then mechanically lifted and incubated with 1 μ g ml⁻¹ Alexa647 (Invitrogen)-conjugated M1 anti-Flag monoclonal antibody (Sigma) at 4 °C for 1 h. Mean fluorescence intensity of 10,000 cells was measured using a FACSCalibur instrument (Becton Dickinson). Each condition was performed in triplicate.

Enhanced GFP calibration. Recombinant eGFP (BioVision) was used for calibrating average fluorescence intensity of the biosensors, imaged in confocal optical sections through the cytoplasm of cells not exposed to agonist (to achieve diffuse cytoplasmic distribution of the biosensors). eGFP was diluted in Hank's balanced salt solution and confocal sections were imaged through droplets of each using the same illumination and acquisition parameters as for imaging the biosensors in cells. For each cell, a background fluorescence value was determined by average fluorescence intensity of a blank region in the same image. The cytoplasmic concentration of biosensors was estimated by interpolation of the background-subtracted value using a linear least-squares fit to the standard plot.

Generation of β_2 -AR-rHDL nanoparticles. Apolipoprotein-AI (Apo-AI) was biotinylated using NHS-PEG4-biotin (Pierce Biotechnology) at a 1:1 molar ratio. Following a 30-min biotinylation reaction at room temperature, the sample was dialysed to remove free biotin. Flag-tagged β_2 -AR was incorporated into recombinant high density lipoprotein (rHDL) particles as previously described^{34,35} using biotinylated Apo-AI. Receptor-containing particles were then purified by M1 anti-Flag immunoaffinity chromatography³⁶. Particles containing purified monobromobimane-labelled β_2 -AR were generated similarly except not using biotinylated HDL, with receptor labelling and fluorescence analysis carried out as described previously¹⁸.

Assessing Nb80 binding to immobilized β_2 AR-rHDL. Nb80 binding to unliganded and agonist-occupied β_2 -AR was measured using the OctetRED biolayer interferometry system (Pall FortéBio). In this assay, a target protein is immobilized on the functionalized tip of a fibre optic probe that is dipped into an analyte solution to observe analyte association to the target protein. A dissociation step is then performed by transferring the biosensor into buffer lacking analyte. Analyte association/dissociation is measured by monitoring changes in the interference pattern of a light beam reflected from the biosensor tip as the total mass bound at the tip surface changes³⁷. Streptavidin-coated biosensors (Pall FortéBio) were loaded with biotinylated β_2 -AR-rHDL particles for 15 min at room temperature and the biosensors were transferred to the OctetRED instrument. Sensors were placed into assay buffer (20 mM HEPES, pH 7.7, 100 mM NaCl, 1 mM EDTA, 0.02% (w/v) ascorbic acid, 0.05% (w/v) BSA) with or without 100 μ M isoprenaline for 30 min. To measure Nb80 association, the sensor was transferred to assay buffer with Nb80 (at indicated concentrations) for 5 min, followed by a 30 min dissociation step in assay buffer. Isoprenaline (100 μ M) was included in the association and dissociation steps when measuring Nb80 binding to agonist-occupied receptor. All experiments were carried out at 25 °C with the assay plate shaking at 1,000 r.p.m. Buffer-only controls were included in each experiment to monitor for baseline drift, and nonspecific Nb80 binding was measured in a parallel assay using sensors loaded with empty rHDL particles. Raw data were

processed to remove baseline and nonspecific binding using Octet Data Analysis 7.0 software (Pall FortéBio) and exported to Prism 5 (GraphPad) for curve fitting. All association and dissociation curves were fit using a single-phase exponential association or decay curves, respectively. Equilibrium binding affinity of Nb80 for β_2 -AR in the presence or absence of the agonist isoprenaline was assessed by monitoring the maximal interference shift generated by Nb80 binding (at varying Nb80 concentrations) to the probe containing β_2 -AR reconstituted in rHDL. The maximal shift was plotted against the Nb80 concentration and fitted by nonlinear regression in Prism 5 (GraphPad) to generate the apparent affinity.

Inhibition of bodipy-GTP γ S-FL binding by Nb37. The effect of Nb37 on GTP loading of purified G proteins was measured using 100 nM bodipy-GTP γ S-FL (Invitrogen) essentially as described²⁵. In this assay we used the fluorescence emission of bodipy-GTP γ S-FL ($\lambda_{\text{ex}} \sim 470$ nm, $\lambda_{\text{em}} \sim 515$ nm) that accompanies binding of the labelled nucleotide to G protein³⁸. Briefly, the fluorescence of 100 nM bodipy-GTP γ S-FL was measured in the presence of 1 mM of the indicated G protein using a 96-well microtitre plate format on a M5 fluorescence plate reader (Molecular Devices). Nb37 was added together with bodipy-GTP γ S-FL and the binding reaction was initiated by the addition of G protein (1 mM) in 20 mM Tris-HCl, pH 8.0, 3 mM MgCl₂, 1 mM dithiothreitol in a final volume of 200 μ l. Bodipy-GTP γ S-FL binding to heterotrimeric G protein included 0.1% dodecyl-maltoside (final). G α_s was purified as described³⁹. G $\alpha_s\beta\gamma$ was purified as described¹³. Myristoylated G α_i was purified as described⁴⁰. The time scans were limited to 240 s to minimize the accumulation of hydrolysis of the product of bodipy-GTP γ S-FL, bodipy-phosphate⁴¹.

31. Gage, R. M., Matveeva, E. A., Whiteheart, S. W. & von Zastrow, M. Type I PDZ ligands are sufficient to promote rapid recycling of G protein-coupled receptors independent of binding to *N*-ethylmaleimide-sensitive factor. *J. Biol. Chem.* **280**, 3305–3313 (2005).
32. Puthenveedu, M. A. *et al.* Sequence-dependent sorting of recycling proteins by actin-stabilized endosomal microdomains. *Cell* **143**, 761–773 (2010).
33. Yudowski, G. A., Puthenveedu, M. A., Henry, A. G. & von Zastrow, M. Cargo-mediated regulation of a rapid Rab4-dependent recycling pathway. *Mol. Biol. Cell* **20**, 2774–2784 (2009).
34. Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
35. Kuszak, A. J. *et al.* Purification and functional reconstitution of monomeric μ -opioid receptors: allosteric modulation of agonist binding by Gi₂. *J. Biol. Chem.* **284**, 26732–26741 (2009).
36. Yao, X. *et al.* Coupling ligand structure to specific conformational switches in the β_2 -adrenoceptor. *Nature Chem. Biol.* **2**, 417–422 (2006).
37. Abdiche, Y., Malashock, D., Pinkerton, A. & Pons, J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Anal. Biochem.* **377**, 209–217 (2008).
38. McEwen, D. P., Gee, K. R., Kang, H. C. & Neubig, R. R. Fluorescent BODIPY-GTP analogs: real-time measurement of nucleotide binding to G proteins. *Anal. Biochem.* **291**, 109–117 (2001).
39. Sunahara, R. K., Tesmer, J. J., Gilman, A. G. & Sprang, S. R. Crystal structure of the adenylyl cyclase activator G α_s . *Science* **278**, 1943–1947 (1997).
40. Lee, E., Linder, M. E. & Gilman, A. G. Expression of G-protein α subunits in *Escherichia coli*. *Methods Enzymol.* **237**, 146–164 (1994).
41. Jameson, E. E. *et al.* Real-time detection of basal and stimulated G protein GTPase activity using fluorescent GTP analogues. *J. Biol. Chem.* **280**, 7712–7719 (2005).

CAREERS

TURNING POINT Virologist follows her dream to become an astronaut **p.541**

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com

IMAGES.COM/CORBIS



PUBLISHING

Open to possibilities

Opting for open access means considering costs, journal prestige and career implications.

BY STEPHEN PINCOCK

Not so long ago, Mike Taylor gave no thought to open access when submitting research papers for publication. “I just sent them to the well-respected journals,” says Taylor, a palaeontologist at the University of Bristol, UK. “It’s only gradually, really, that it became apparent to me how stupid that was.”

As he started to think more carefully about where he published his work, his contemplation turned into fully fledged advocacy. Taylor would like to see free, unrestricted access to all scholarly papers online. In an article on the website of UK newspaper *The Guardian* this

January, he argued that hiding publications behind a paywall is “immoral” (see go.nature.com/v9fmtm). But within hours of appearing, the article was accumulating comments that laid out caveats that he had not considered.

The point that raised the most ire was Taylor’s argument that scientists no longer need to publish in prestigious journals to boost their careers, which commenters vigorously refuted. Taylor was forced to rethink his position on

that: journal prestige remains important, he realized. “I got quite a lot of criticism from people I respect a lot,” he says.

Early-career scientists face a pressing question: how should they publish to advance their careers at a time when the scholarly world is being shaken up? New options in publishing have highlighted sticky debates related to the impacts of costs on individual researchers and the ethics of business models that can keep cash-poor scientists from accessing data. But many researchers contend that impact factors and other metrics of journal prestige remain crucial — for now.

Over the past 20 years, open-access ►



► publishing has become a major part of the scholarly landscape. It is now common in astronomy, maths and physics, where most researchers submit their work to the open-access repository arXiv.org before it is published, and is on the rise in the life sciences and other fields. Over the past decade, open-access publishing has increased its share of articles by about 1% a year (see page 425). Around 17% of the 1.66 million articles indexed by abstract and citation database Scopus in 2011 were freely available from journal publishers¹.

Worldwide, more than 200 institutions and 80 research funders require their researchers' work to be open access, according to the Roar-map registry (roarmap.eprints.org). For example, from 1 April, researchers supported by any of the seven UK research councils will be asked to publish their work in a journal that either provides immediate and unrestricted access to the final published version of the paper, or consents to the manuscript being deposited in an open-access repository within a certain time — six months for science papers. The US National Institutes of Health requires that scientists submit final peer-reviewed journal manuscripts arising from agency-funded work to the digital archive PubMed Central, and that those papers are made available to the public within one year of publication.

THE POWER OF PRESTIGE

Not everyone shares Taylor's moral outrage over the need for open access. Many senior researchers have simple advice, especially for early-career scientists: go to the best journal you can publish in. Rob Brooks, an evolutionary scientist at the University of New South Wales in Sydney, Australia, supports open access in principle, but says that career building still relies on established models of prestige. "Journal quality remains the benchmark for that piece of work and that's what people will be assessed by," he says. "Impact factors still pretty much rule. A lot of people — grant committees, administrators and even referees — can't assess quality. All they can do is count or pseudo-quantify. They count the number of papers you've got and count the impact factors of the papers and make a little metric, rather than just reading the papers."

Early-career scientists should be wary about turning away from high-impact journals to publish in open-access outlets with lesser reputations, says Chris Chambers, a neuroscientist at Cardiff University, UK, and an academic editor at *PLoS ONE*, who was among those who critiqued Taylor's post. "On the one hand, this is a noble act that is impossible not to applaud," he says. "On the other hand, there is a risk that such researchers deal themselves out of the game, being overtaken in the race to senior positions by more careerist colleagues." But going for high-impact journals need not rule out open-access, says Peter Suber, director of the open-access project at Harvard University in Cambridge, Massachusetts. He suggests

looking through the Directory of Open Access Journals (www.doaj.org) for those with editors and published authors who are familiar and respected in the field. Suber also points out that publishing in a subscription journal does not mean that an article cannot be made freely available online later; most allow pre- or post-print archiving in open-access repositories such as those listed at www.orphn.org.

The open-access journal *eLife* published its first articles only last October, but can already boast a measure of prestige. Although it is too early to gauge its impact factor, the journal seems likely to benefit from the reputation of its founders, research funders the Wellcome Trust in London, the Howard Hughes Medical Institute in Chevy Chase, Maryland, and the Max Planck Society in Munich, Germany. "*eLife* is ... an obvious example of a brand-new journal with no branding, no standing in the world," says Peter Binfield, publisher of the separate open-access journal *PeerJ*, based in San Francisco, California, "and yet it's clear that this is a good place to publish, because it's got some big-name backers."



"If you come to the Wellcome for a grant, funding decisions are based on the intrinsic merit of the work, and not the title of the journal."

Robert Kiley

published." Kiley points to the policies of the UK programme for assessing research quality, the Research Excellence Framework, which stated in July 2012 that no grant-review sub-panel "will make any use of journal impact factors, rankings, lists or the perceived standing of publishers in assessing the quality of research outputs" (see go.nature.com/xgsses).

In principle, open-access publishing can connect researchers to a wider readership. Some studies suggest that articles in open-access journals or repositories are accessed more often and reach a broader audience than those in subscription-only journals. Whether this translates into higher citation rates is up for debate. In 2010, a meta-analysis found 27 studies showing that open-access articles had more citations than papers behind paywalls — up to 600% more, depending on the field — and

four that found no open-access advantage².

Some journals, including *PLoS ONE* and *PeerJ*, tally metrics such as the number of people who have clicked on each paper, and the paper's visibility on social-media websites such as Twitter and Facebook. Ross Mounce, a palaeontology PhD student at the University of Bath, UK, thinks that such alternative metrics are already beginning to free researchers from conventional measures of prestige. "I think if you have faith in your own work, then you can publish it wherever," he says, countering comments on Taylor's article. "And as long as people can access it and read about it and know about it, and can discover your work, then that's all you need."

WEIGHING THE COST

Many researchers contemplating open access are concerned about having to pay publication costs. As of August 2011, about one-quarter of open-access journals said that they charged article-processing fees; in 2010, those fees ranged between US\$8 and \$3,900, with the average around \$900 (ref. 3).

According to Carl Bergstrom, an evolutionary biologist at the University of Washington in Seattle, there tends to be a positive correlation between an open-access journal's fees and its score in a system he co-developed that rates journals according to the number of citations they receive, with citations from highly ranked journals weighted more heavily. However, not all expensive journals have high ratings; an interactive tool shows that there is a great deal of variation (www.eigenfactor.org/openaccess).

In any case, Suber notes that only a little over 10% of article-processing fees are paid out of pocket by authors⁴. "Most of the time, those fees are paid on behalf of the author by the author's employer or the author's funding agency," he says. Scientists should check with their funding agencies directly, but open-access publisher BioMed Central keeps an incomplete list of those that will pay (see go.nature.com/pqjxoi). And some publishers reduce or waive fees for researchers who cannot afford to pay.

PeerJ, which published its first papers last month, has taken a completely new approach to cost. Rather than charging a fee per paper, it offers lifetime membership, ranging from \$99 for one publication per year to \$299 for unlimited publications (see *Nature* 486, 166; 2012). All memberships include the right to post articles on a non-peer-reviewed preprint server. For a paper to be published, all authors must be members (or at least 12 authors must be, for papers with more). "We've deliberately tried to strip out all the extraneous costs from the system to make it as efficient and cheap as possible from a researcher's point of view," says Binfield. "Then we've layered on this different business model, which is a payment per person per membership rather than payment per publication."

Bo-Christer Björk, an information systems scientist at Hanken School of Economics in

Helsinki, says that the number of subscription journals offering open access for a fee has doubled in recent years, and currently stands at more than 4,300. However, just 1–2% of eligible authors take up that open-access option⁵. Björk argues that the high costs are a deterrent — about \$3,000 is typical. “There are, however, a few publishers and individual journals with a much higher uptake,” he says. For example, *Nature Communications*, which was launched in 2010 by Nature Publishing Group (NPG), is a multidisciplinary journal that takes this hybrid approach. When it started, about half of authors chose the open-access option. But those numbers have fluctuated, says James Butcher, associate director of open publishing at NPG, and in the past six months about 30% have chosen open access. Butcher has not collected data to explain the trend, but speculates that it might be attributable to the journal’s relatively high fees of around \$5,000, to changing author demographics or to a general drop in interest after early excitement about the open-access option.

Chambers suggests that paying a fee to publish in a hybrid journal is a good way to achieve both accessibility and prestige. “The key factor is whether the researcher has the funds available to do so, and whether the funding agency requires papers to be open access,” he says. But where funds have not been set aside for article-processing charges, and young scientists would need to spend research money, Chambers advises against consenting to high publishing fees.

“As a young investigator you have to do what’s economically viable,” says Stephen Macknik, a neuroscientist at the Barrow Neurological Institute in Phoenix, Arizona. Paying an article-processing charge for a reputable open-access journal may be a good middle ground for young researchers, he says.

But scientists shouldn’t sacrifice funding that was meant for research. “To maximize their competitiveness it is vital that young researchers maintain a productive profile of high-quality research, and this means using research funds to do as much high-quality research as possible,” says Chambers. “It falls to the more senior scientists to change the system.” ■

Stephen Pincock is a freelance writer based in Sydney, Australia.

1. Laakso, M. & Björk, B.-C. *BMC Med.* **10**, 124 (2012).
2. Swan, A. *The Open Access Citation Advantage: Studies and Results to Date*. (Univ. Southampton, 2010); available at go.nature.com/8xg751
3. Solomon, D. J. & Björk, B.-C. *J. Am. Soc. Inform. Sci. Tech.* **63**, 1485–1495 (2012).
4. Dallmeier-Tiessen, S. *et al.* Preprint available at <http://arxiv.org/abs/1101.5260> (2011).
5. Björk, B.-C. *J. Am. Soc. Inform. Sci. Tech.* **63**, 1496–1504 (2012).

TURNING POINT

Kate Rubins

After 15 years in infectious-disease biology, Kate Rubins jumped at the chance to fulfil her childhood dream of becoming an astronaut.

Did you always want to become an astronaut?

As a kid, I really did, but various people pointed out that it was not the most realistic career choice. When I was 16, my dad took me to a DNA conference at the Exploratorium science museum in San Francisco, California, and I was captivated by this way of looking at biology and by the discussions of bits of nucleic acid that could make us sick.

How did you come to focus on research related to public health?

As an undergraduate majoring in biology at the University of California, San Diego, I worked on infectious diseases at the nearby Salk Institute for Biological Studies. I decided to do graduate studies in virology at Stanford University in California because it had a hospital, which made working on clinical applications easier. I was looking at immune responses related to smallpox and Ebola, so I flew to Maryland every few weeks to work in a biosafety-level-4 lab, which handles the most dangerous microbes. Then I shipped the data back to Stanford.

You built a lab quickly after your PhD. How?

I decided to skip the postdoc. The Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, had a fellows programme that was akin to a junior faculty position with few teaching responsibilities. That seemed to be a good fit. My interests had shifted to the genomics of infectious disease, and I started laying the groundwork to study monkeypox infections in the Democratic Republic of the Congo. With a lot of hubris, I started my own lab. It was amazing — the Whitehead gave fellows a lot of leeway. In three years, I had secured enough money from the US National Institutes of Health and the departments of defence and homeland security to increase the lab to 14 people.

Why did you apply to become an astronaut?

It was one of those childhood dreams that I couldn’t let go of. I thought that NASA didn’t take biologists and so nothing would come of it, but I knew I would regret it if I did not apply.

How did it feel to have to dismantle your lab after NASA accepted you?

Joining NASA was very exciting, but it was the hardest decision I have had to make in my life. I had been working towards one goal for



NASA

more than 15 years. I had been very specific about what I wanted to do with my career, and this was completely different. I was concerned about my lab members— wonderful people whom I couldn’t leave high and dry. I wanted to make sure that they were able to continue their research. In the six weeks that I was given to shut down my lab before heading to the Johnson Space Center in Houston, Texas, for training, I found good labs for everyone to join.

Three years later, do you still feel that joining NASA was the right decision?

Yes; I am really happy. I have learned a whole bunch of new skills, including how to speak Russian, conduct a space walk and fly a supersonic jet. The whole time at NASA has been a huge turning point for me.

Are you able to do research?

Yes, but it is different from having my own lab. I don’t get to say, “I want to do this”; they select the best peer-reviewed research. I bring an operational perspective to the experiments that match up with my expertise. My time in biosafety labs taught me to work in a high-pressure environment, which provided skills that I am using at NASA. I am working on experiments from immunology to bone loss in microgravity.

Will you go into space?

Fingers crossed. I am in the newest class of astronauts. The International Space Station will be operational until 2020, and perhaps 2028, so there is a chance. NASA is also building a space-launch system to go beyond low-Earth orbit. Whether either of those overlaps with my time frame is unknown, but it would be fantastic. I will go wherever NASA sends me. ■

INTERVIEW BY VIRGINIA GEWIN

STICKY

Go with the flow.

BY GEORGE ZEBROWSKI

Unable to reach Felix by phone, Bruno hobbled across the street to his physicist friend's house and waved the insulting cheque at him. "What's the joke, Felix?"

Long retired, Bruno still recalled the joys of plumbing: the craft of figuring pipes, often well beyond necessity, by eye alone, in quality iron, copper or even plastic connections. It had been a while since anyone had asked him to do repairs, or even make a job estimate. But now a cheque had arrived, with a post-it "thank you" from his lifelong friend attached to the return invoice.

Felix gave him a confused look. "But ... but you did the job while I was away."

"I did not," Bruno said.

"Have you ... forgotten?" Felix asked the sudden stranger.

"I didn't send this invoice," Bruno said.

"I'll see about it," Felix said coldly, closing the door.

A month later, the scene played out again at Felix's door, with Bruno waving invoice and cheque for bathtub fixtures he had not installed.

A third time, Felix said more warmly, "No denying these invoices, with our signatures on each. So who's joshing us, Bruno?"

"Maybe it's you," Bruno said.

"Not you, Bruno?" Felix asked. "Did we both forget?"

Bruno rubbed his bad knee and leaned back on the sofa as they stared at the paperwork on the coffee table in front of them.

"You know, Bruno," Felix said, "I'm only a so-so physicist, but I have a wild idea about all this."

"Really?"

"You're slipping, Bruno," he said.

"Slipping?"

"Through your sum over histories."

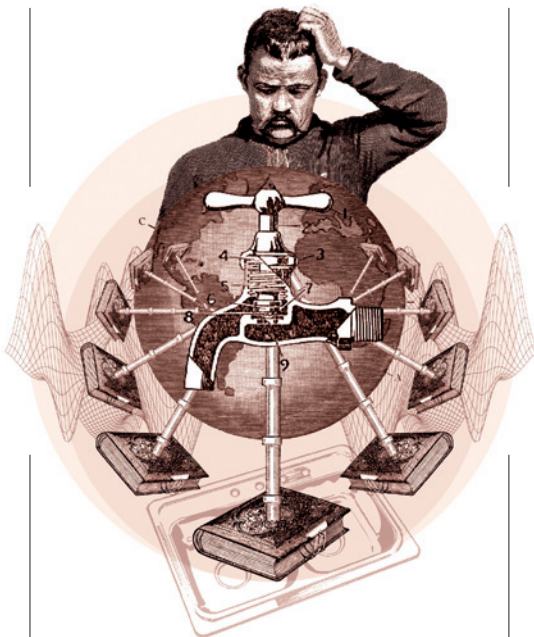
"My what?"

"Across the probabilities. Some version of you did all this work."

"And you paid me," Bruno said, saddened by his friend's covering up what might be signs of delusion and mental decline.

"We can check it out," Felix said.

They went to his bathroom and looked under the sink, where the new drain and water connections glistened with prideful workmanship;



then at the immaculate workmanship in the shower and bath. Felix leaned down under the sink and fondled the drain and water-supply connections. "Ah, yes, nice and dry," he said happily, looking around at the floor. "But look here," he said suddenly, and picked up a long grey hair from the floor. "I'm sure this is yours!" He held the hair up to the light as he stood up and said, "You know, Bruno, if this proves out, you'll be famous in physics, more than I could ever be. It's classic Feynman and Schrödinger."

"Is one hair any proof?" asked Bruno.

"One is enough ... and I'll do the mathematics," Felix said.

"Maybe there'll be another invoice," said Bruno warily.

Felix sighed. "Yes, a paper with all the right maths and one hair won't do. It could have got here a number of ways, and the rest is words. We'll see if it happens again, but even that won't be enough."

"There's not much we can do at the moment," Bruno said, hoping that a delay might rid his overworked friend of delusions. "Moments ..." Felix mused with a glint in his eyes. "Think of all the infinite diverging moments, each one going its own way at every juncture, every minute of the day, at every fraction of a second, throughout all the galaxies!"

"But why — why would all this be happening ... to us?" Bruno asked, humouring his friend.

"Why does anything happen in the quantum realm? No one knows why, except that it does and can be described, even predicted and made useful in technologies. To ask why seems a useless question."

Bruno asked: "But why are we able to notice it?"

"A slip of some kind," Felix said.

"But this ... this spilling up from the small to the large ..."

"The small is affecting the large."

"Are these divergences common?"

Bruno asked.

"They might be finite sets."

Demoralized, Bruno went home and slept badly, dreaming of the objects in his friend's realm of work — a vast grey infinity of spaceless distances and winking points far removed from the undersides of sinks. He woke up thinking that the one domain could have nothing to do with the other, despite the fact that the visible world depended on an invisible realm.

He got up and worked on the page proofs of the new edition of his physics text, which suddenly looked unfamiliar, as if someone else had written it; but finally he settled into its long-worked-over familiarities. "Probabilities are like the thin pages of an infinite book," he wrote in the margin of the manuscript, "and sometimes they stick together, even flow into each other. This may very well be the reality that makes human choices possible, escaping a hard determinism through a form of limited chaos ..."

Clever thought, he told himself.

Sticky stuff.

Some world pages stuck, others didn't. Smiling, he picked up a book and riffled the pages, imagining plumbing analogies to the structure of the Universe ...

There was an annoying knock on his door, and Bruno resented the early interruption.

He opened the door, and saw Felix, his old school friend, with toolbox in hand.

"Well, here I am," Felix said with a smile, looking happy. "Hope you didn't break your kitchen sink just to get me to visit."

Bruno was puzzled for an instant, then stepped aside as the plumber hobbled in. ■

George Zebrowski is an award winning novel and story writer, essayist, poet, editor and lecturer. His latest novel is Empties (Golden Gryphon Press), and with Gregory Benford he is editor of Sentinels In Honor of Arthur C. Clarke (Hadley Rille Books).

Is the ocean food provision index biased?

ARISING FROM B. S. Halpern *et al.* *Nature* **488**, 615–620 (2012)

How close to maximum sustainable food provision is current seafood harvest from the world's oceans? Halpern *et al.*¹ suggest that the answer is 25% from a global index of food provision, part of their multifaceted index of ocean health. Rigorous methods used for management, however, demonstrate that their food provision index is uncorrelated with actual food provision, and that global ocean food provision is in the range of 71–95%. Their results stem from an uncertain method of estimating maximum sustainable yield (MSY), and we believe that this approach should be avoided as a measure of food provision. There is a Reply to this Brief Communication Arising by Halpern, B. S. *et al.* *Nature* **495**, <http://dx.doi.org/10.1038/nature11975> (2013).

The first indication that the Halpern *et al.*¹ food provision index may be unreliable comes from examining the scores of individual countries whose fisheries performance is well understood. It is widely acknowledged that fisheries in New Zealand and the USA are well managed^{2,3} whereas those in Europe are not^{3,4}. However, New Zealand has an index value of 16% and the USA 25%, whereas Europe includes four of the six highest values, topped by France (87%) and Spain (60%).

We tested the reliability of their index in a number of ways. Crudely, their index should be about the same as dividing global catches by global MSY. Recent estimates of global MSY are 83–99 million t (ref. 5), 89.5–100.9 million t (ref. 6) and 85–110 million t (ref. 7), which are similar to older estimates that are often rounded to 100 million t (refs 8, 9). Dividing current catch of 78.6 million t (ref. 10) by MSY yields estimates for current food provision of 71–95%.

The underlying basis for the Halpern *et al.*¹ food provision index (setting aside modifications for trends, pressures and taxonomic

groupings) involves comparing current catches to estimated multi-species MSY as estimated from a regression of MSY against maximum catch (see Fig. 1 for details). This regression is highly uncertain: the ratio of their predicted MSY to actual MSY spans a 50-fold range (95% CI, 0.10–5.02). This is why MSY is normally estimated from biomass data using stock assessments. Stock assessment estimates can be used to ground-truth the Halpern *et al.* catch-based estimates of MSY. For example, when their method is applied to all assessed tuna and billfish fisheries, their predicted MSY is lower than assessment MSY values for the 29 (out of 36) largest fisheries, and summed predicted MSY is only 41% of summed biomass-estimated MSY. When these estimates are fed into their rule for calculating food provision (which is zero when catches exceed $1.5 \times$ MSY), their food provision index for tuna and billfish species is 0%. By contrast, the MSY estimates from the assessments used to manage these tuna and billfish species result in an estimate of food provision of 94% (summed catches divided by summed MSY).

For a more rigorous examination, we compared sustainable yield at a given biomass level to MSY, both from a Pella–Tomlinson model (Methods). We applied this model to a compilation of 234 fisheries stock assessments^{11,12} for which we have reliable estimates of catch, biomass and MSY. For these well-studied fisheries, covering 20–25% of global catch, food provision is 86% (median), or 75% (weighted by maximum catch). Notably, food provision estimates from these fisheries are uncorrelated with the Halpern *et al.*¹ estimates (Fig. 2,

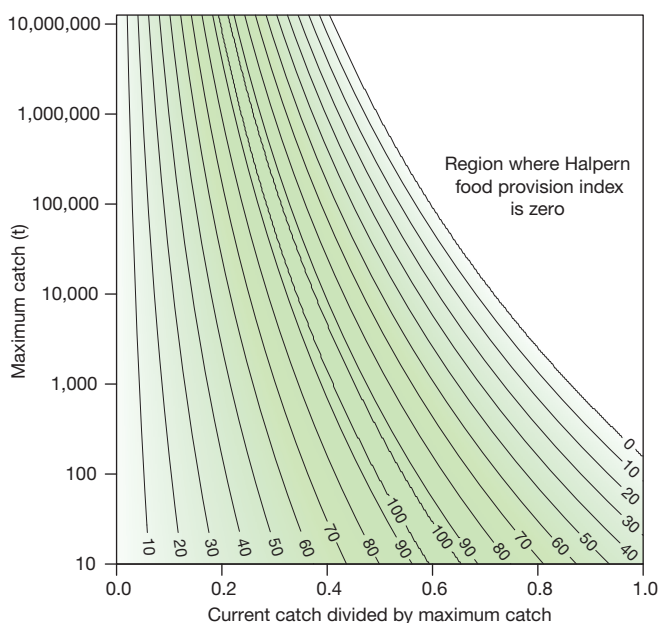


Figure 1 | Contours of the Halpern *et al.*¹ food provision index (0–100%) in relation to current catch and maximum catch (C_{\max}). The Halpern index is 100% when catch is between 71% and 79% of assumed MSY, and 0% when catch is either zero or greater than 150% of assumed MSY. MSY is assumed to equal $C_{\max}^{0.92}$, based on a highly uncertain relation between MSY and C_{\max} . For large fisheries, the food provision index is zero whenever current catches are more than 50% of maximum catches.

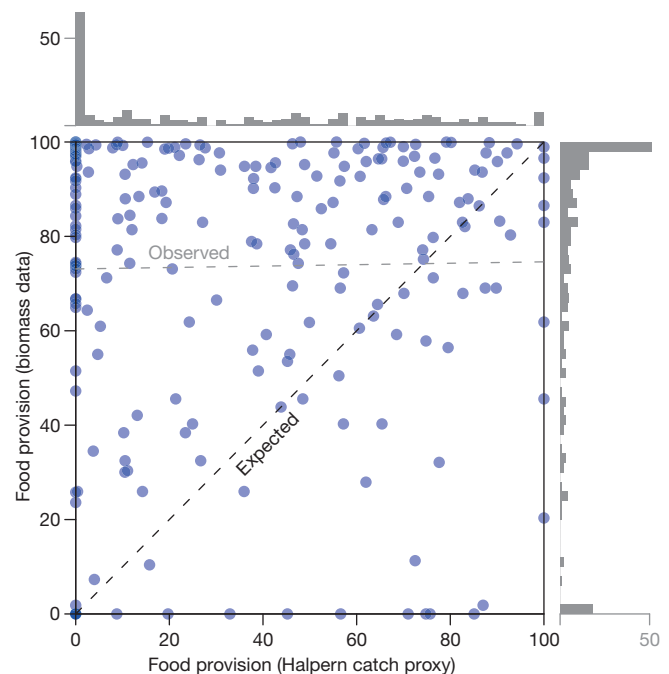


Figure 2 | Food provision indices (percentages) estimated using the Halpern *et al.*¹ proxy, compared to food provision based on biomass data for 234 fisheries comprising 20–25% of global catch (circles, see Methods for details). The two indices should be tightly correlated and fall on the 1:1 'expected' line, but are uncorrelated ('observed' line, $y = 0.015x + 73.1$, $R^2 = 0.0003$). Grey histograms indicate the total number of fisheries at each value for the food provision values, with a peak at 0% for the Halpern *et al.*¹ proxy and at 100% for the index based on biomass data.

$R^2 = 0.0003$). Furthermore, for fisheries with a food provision index of 0% from Halpern *et al.*¹, actual food provision averages 73%. As the Halpern food provision index is uncorrelated with food provision estimates based on well-studied fisheries in many regions of the world¹¹, we see little value in applying it to data-poor fisheries on a country-by-country basis.

The same rigorous method can be applied to Food and Agriculture Organization of the United Nations (FAO) status reports¹⁰, which covered 70% of global catches in 2009. Re-sampling fisheries from FAO ranges for overexploited fisheries, fully exploited fisheries and underexploited fisheries³ and estimating food provision using the Pella–Tomlinson model yields a mean food provision index of 82% (95% CI, 81–84%).

Thus, food provision from oceanic fisheries is consistently estimated to be in the range of 71–95% by a wide variety of methods, which is far from the 25% estimated by Halpern *et al.*¹. In conclusion, Halpern *et al.*¹ believe that it must be possible to extract MSY and food provision estimates from a simple function of current catch and maximum catch. Such approaches in other guises are similarly flawed^{12–15}.

Methods

Sustainable yield from fisheries as a function of population size can be represented by a Pella–Tomlinson curve¹⁶. From this curve, food provision is the ratio of sustainable yield to MSY, and can be written as a function of current biomass (B) relative to biomass producing MSY (B_{MSY}) plus a parameter (z) that governs the shape of the curve:

food provision =

$$100 \times \max \left\{ 0, \frac{z^{z/(z-1)}}{z-1} \left(\left(\frac{1}{z} \right)^{1/(z-1)} \frac{B}{B_{\text{MSY}}} - \left[\left(\frac{1}{z} \right)^{1/(z-1)} \frac{B}{B_{\text{MSY}}} \right]^z \right) \right\}$$

Setting $z = 1.188$ ensures that B_{MSY} occurs at 40% of unexploited biomass, which is the mean from 147 fisheries¹⁶. This food provision index is 0% when biomass is zero or greater than $2.5 \times B_{\text{MSY}}$, and 100% when biomass equals B_{MSY} .

This project was partly funded by Cooperative Agreement NA09OAR4320075 between the Joint Institute for Marine and Atmospheric Research (JIMAR) and the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not necessarily reflect the views of the NOAA or any of its subdivisions.

Trevor A. Branch¹, Daniel J. Hively¹ & Ray Hilborn¹

¹School of Aquatic & Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195, USA.
email: tbranch@uw.edu

Received 31 October 2012; accepted 28 January 2013.

1. Halpern, B. S. *et al.* An index to assess the health and benefits of the global ocean. *Nature* **488**, 615–620 (2012).
2. Pauly, D. *Focusing One's Microscope: Daniel Pauly on the State of Global Fisheries* <http://blog.nature.org/2011/01/daniel-pauly-fish-stock-global-world-fisheries/> (2011).
3. Worm, B. *et al.* Rebuilding global fisheries. *Science* **325**, 578–585 (2009).
4. Froese, R. & Proelß, A. Rebuilding fish stocks no later than 2015: will Europe meet the deadline? *Fish Fish.* **11**, 194–202 (2010).
5. Sumaila, U. R. *et al.* Benefits of rebuilding global marine fisheries outweigh costs. *PLoS ONE* **7**, e40542 (2012).
6. Arnason, R., Kelleher, K. & Willmann, R. *The Sunken Billions. The Economic Justification for Fisheries Reform.* (World Bank and FAO, 2008).
7. Costello, C. *et al.* Status and solutions for the world's unassessed fisheries. *Science* **338**, 517–520 (2012).
8. Ryther, J. Photosynthesis and fish production in the sea. *Science* **166**, 72–76 (1969).
9. Pauly, D. One hundred million tonnes of fish, and fisheries research. *Fish. Res.* **25**, 25–38 (1996).
10. FAO. *The State of World Fisheries and Aquaculture 2012* (Food and Agriculture Organization of the United Nations, 2012).
11. Ricard, D., Minto, C., Baum, J. K. & Jensen, O. P. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish Fish.* **13**, 380–398 (2012).
12. Branch, T. A., Jensen, O. P., Ricard, D., Ye, Y. & Hilborn, R. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. *Conserv. Biol.* **25**, 777–786 (2011).
13. Daan, N., Gislason, H., Pope, J. G. & Rice, J. C. Apocalypse in world fisheries? The reports of their death are greatly exaggerated. *ICES J. Mar. Sci.* **68**, 1375–1378 (2011).
14. Carruthers, T. R., Walters, C. J. & McAllister, M. K. Evaluating methods that classify fisheries stock status using only fisheries catch data. *Fish. Res.* **119–120**, 66–79 (2012).
15. Hilborn, R. & Branch, T. A. Does catch reflect abundance? No, it is misleading. *Nature* **494**, 303–306 (2013).
16. Thorson, J. T., Cope, J. M., Branch, T. A. & Jensen, O. P. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Can. J. Fish. Aquat. Sci.* **69**, 1556–1568 (2012).

Author Contributions T.A.B. designed the study, T.A.B. and D.J.H. analysed the data, and T.A.B. and R.H. discussed the results and wrote the paper together.

Competing Financial Interests Declared none.

doi:10.1038/nature11974

Halpern *et al.* replyREPLYING TO T. A. Branch, D. J. Hively & R. Hilborn. *Nature* **495**, <http://dx.doi.org/10.1038/nature11974> (2013)

Branch *et al.*¹ suggest that the fisheries component of our ocean health and benefits index² is 'biased'. We contend that our approach is no more biased than their alternative, and that our method improves on theirs in three fundamental ways: it provides a score for every country, accounts for thousands of data-poor stocks, and incorporates sustainability more comprehensively.

Our global score is an area-weighted country average of a multi-dimensional assessment of sustainable food provision, fundamentally different from what Branch *et al.*¹ interpret it to be, which undermines comparisons to their global estimate. They incorrectly state that our index should be roughly equivalent to dividing global catch by global maximum sustainable yield (MSY), and incorrectly interpret a score of 100 in our index as $B = B_{\text{MSY}}$. They focus on single-stock biomass harvested relative to single-stock MSY, as in their Fig. 1, and assume that they know true MSY values. We used a multispecies reference point (mMSY) with a precautionary buffer that penalizes countries both above and below this value, and our focus was on fisheries within exclusive economic zone (EEZ) scales. Branch *et al.*¹ focused on a single global assessment (of current status only), leading to a relatively small number of the largest fish stocks worldwide driving their interpretation. We chose to use catch data, available for all countries and most stocks, acknowledging the associated limitations and uncertainty of those data. Although their approach may be useful for assessments of global biomass production, and we agree it is preferable to use abundance estimates and other indicators rather than catch data when possible, reliance on such data precludes country-level assessments at the global scale.

By focusing their analysis on a geographically restricted data set of exclusively large commercial species (<5% of all stocks in the world), they assume that small stocks, and the many interactions they have with large stocks, do not matter for understanding sustainability, and that the many data-poor stocks are irrelevant to understanding food provision. Data-poor stocks represent ~80% of all fisheries biomass and the vast majority of stocks in most countries. These stocks are in worse condition than assessed stocks, and getting worse³. Their selection of tuna and billfish as a counter example to our results focuses on an outlier; these large pelagic stocks are not representative of most fisheries within EEZs (that is, bottom-dwelling species or those taken in small-scale fisheries) or most unassessed stocks. Therefore, it is reasonable to assume that unassessed stocks would lower fisheries index scores, and for our comprehensive assessment to differ from the narrower one of Branch *et al.*¹

Because of these key differences, strong correlations between our methods should not necessarily be expected. In addition, assessments

made using best available data in individual countries should generate different outcomes compared with a global assessment where the focus is on using globally available and comparable data. We know that many country-specific scores will change as we improve the underlying data and our ability to predict sustainable yield of stocks around the world. Our index was explicitly designed as a framework that can ingest the best scientific information as it becomes available, whether as better data or improved approaches to measuring a goal, and in fact, we are doing just that as we apply the index at regional scales. Authors from both papers are currently collaborating with the Food and Agriculture Organization of the United Nations to examine how to improve methodologies for estimating fisheries exploitation status globally. We look forward to incorporating those results and improved data in future iterations of the index.

This Reply is written by the subset of authors that developed the fisheries component of the ocean health and benefits index² relevant to the discussion in Branch *et al.*¹.

Benjamin S. Halpern^{1,2}, Steven D. Gaines³, Kristin Kleisner⁴, Catherine Longo¹, Daniel Pauly⁴, Andrew A. Rosenberg⁵, Jameal F. Samhouri⁶ & Dirk Zeller⁴

¹National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA.

²Center for Marine Assessment and Planning, University of California, Santa Barbara, California 93106, USA.
email: halpern@nceas.ucsb.edu

³Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, USA.

⁴Sea Around Us Project, Fisheries Centre, 2202 Main Hall, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

⁵Union of Concerned Scientists, 2 Brattle Square, Cambridge, Massachusetts 02138, USA.

⁶Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, Washington 98112-2097, USA.

1. Branch, T. A., Hively, D. J. & Hilborn, R. Is the ocean food provision index biased? *Nature* **495**, <http://dx.doi.org/10.1038/nature11974> (2013).
2. Halpern, B. S. *et al.* An index to assess the health and benefits of the global ocean. *Nature* **488**, 615–620 (2012).
3. Costello, C. *et al.* Status and solutions for the world's unassessed fisheries. *Science* **338**, 517–520 (2012).

doi:10.1038/nature11975

Is the ocean food provision index biased?

ARISING FROM B. S. Halpern *et al.* *Nature* **488**, 615–620 (2012)

How close to maximum sustainable food provision is current seafood harvest from the world's oceans? Halpern *et al.*¹ suggest that the answer is 25% from a global index of food provision, part of their multifaceted index of ocean health. Rigorous methods used for management, however, demonstrate that their food provision index is uncorrelated with actual food provision, and that global ocean food provision is in the range of 71–95%. Their results stem from an uncertain method of estimating maximum sustainable yield (MSY), and we believe that this approach should be avoided as a measure of food provision. There is a Reply to this Brief Communication Arising by Halpern, B. S. *et al.* *Nature* **495**, <http://dx.doi.org/10.1038/nature11975> (2013).

The first indication that the Halpern *et al.*¹ food provision index may be unreliable comes from examining the scores of individual countries whose fisheries performance is well understood. It is widely acknowledged that fisheries in New Zealand and the USA are well managed^{2,3} whereas those in Europe are not^{3,4}. However, New Zealand has an index value of 16% and the USA 25%, whereas Europe includes four of the six highest values, topped by France (87%) and Spain (60%).

We tested the reliability of their index in a number of ways. Crudely, their index should be about the same as dividing global catches by global MSY. Recent estimates of global MSY are 83–99 million t (ref. 5), 89.5–100.9 million t (ref. 6) and 85–110 million t (ref. 7), which are similar to older estimates that are often rounded to 100 million t (refs 8, 9). Dividing current catch of 78.6 million t (ref. 10) by MSY yields estimates for current food provision of 71–95%.

The underlying basis for the Halpern *et al.*¹ food provision index (setting aside modifications for trends, pressures and taxonomic

groupings) involves comparing current catches to estimated multi-species MSY as estimated from a regression of MSY against maximum catch (see Fig. 1 for details). This regression is highly uncertain: the ratio of their predicted MSY to actual MSY spans a 50-fold range (95% CI, 0.10–5.02). This is why MSY is normally estimated from biomass data using stock assessments. Stock assessment estimates can be used to ground-truth the Halpern *et al.* catch-based estimates of MSY. For example, when their method is applied to all assessed tuna and billfish fisheries, their predicted MSY is lower than assessment MSY values for the 29 (out of 36) largest fisheries, and summed predicted MSY is only 41% of summed biomass-estimated MSY. When these estimates are fed into their rule for calculating food provision (which is zero when catches exceed $1.5 \times$ MSY), their food provision index for tuna and billfish species is 0%. By contrast, the MSY estimates from the assessments used to manage these tuna and billfish species result in an estimate of food provision of 94% (summed catches divided by summed MSY).

For a more rigorous examination, we compared sustainable yield at a given biomass level to MSY, both from a Pella–Tomlinson model (Methods). We applied this model to a compilation of 234 fisheries stock assessments^{11,12} for which we have reliable estimates of catch, biomass and MSY. For these well-studied fisheries, covering 20–25% of global catch, food provision is 86% (median), or 75% (weighted by maximum catch). Notably, food provision estimates from these fisheries are uncorrelated with the Halpern *et al.*¹ estimates (Fig. 2,

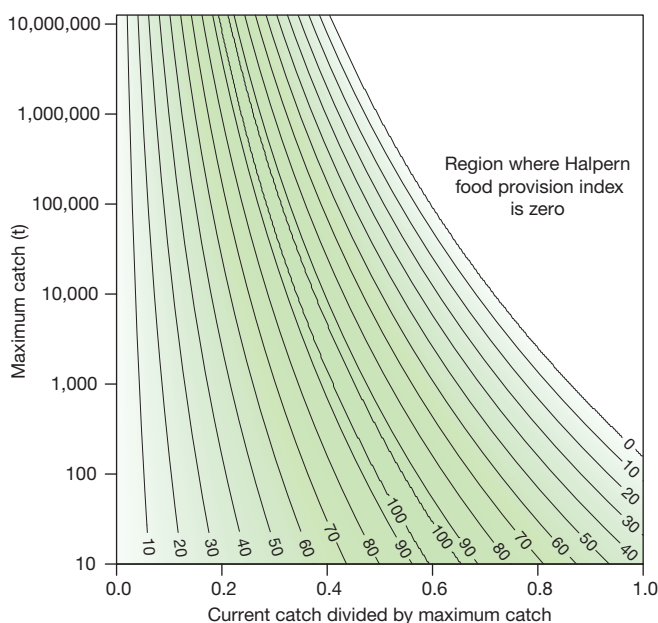


Figure 1 | Contours of the Halpern *et al.*¹ food provision index (0–100%) in relation to current catch and maximum catch (C_{\max}). The Halpern index is 100% when catch is between 71% and 79% of assumed MSY, and 0% when catch is either zero or greater than 150% of assumed MSY. MSY is assumed to equal $C_{\max}^{0.92}$, based on a highly uncertain relation between MSY and C_{\max} . For large fisheries, the food provision index is zero whenever current catches are more than 50% of maximum catches.

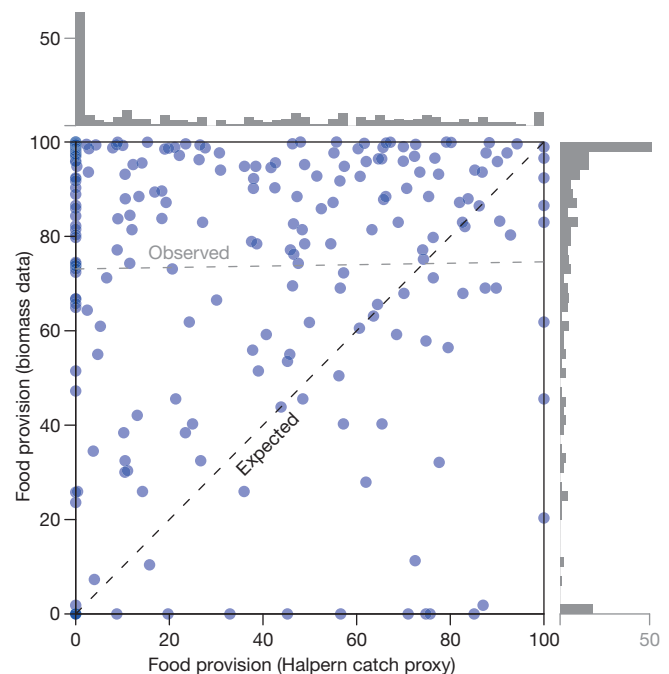


Figure 2 | Food provision indices (percentages) estimated using the Halpern *et al.*¹ proxy, compared to food provision based on biomass data for 234 fisheries comprising 20–25% of global catch (circles, see Methods for details). The two indices should be tightly correlated and fall on the 1:1 'expected' line, but are uncorrelated ('observed' line, $y = 0.015x + 73.1$, $R^2 = 0.0003$). Grey histograms indicate the total number of fisheries at each value for the food provision values, with a peak at 0% for the Halpern *et al.*¹ proxy and at 100% for the index based on biomass data.

$R^2 = 0.0003$). Furthermore, for fisheries with a food provision index of 0% from Halpern *et al.*¹, actual food provision averages 73%. As the Halpern food provision index is uncorrelated with food provision estimates based on well-studied fisheries in many regions of the world¹¹, we see little value in applying it to data-poor fisheries on a country-by-country basis.

The same rigorous method can be applied to Food and Agriculture Organization of the United Nations (FAO) status reports¹⁰, which covered 70% of global catches in 2009. Re-sampling fisheries from FAO ranges for overexploited fisheries, fully exploited fisheries and underexploited fisheries³ and estimating food provision using the Pella–Tomlinson model yields a mean food provision index of 82% (95% CI, 81–84%).

Thus, food provision from oceanic fisheries is consistently estimated to be in the range of 71–95% by a wide variety of methods, which is far from the 25% estimated by Halpern *et al.*¹. In conclusion, Halpern *et al.*¹ believe that it must be possible to extract MSY and food provision estimates from a simple function of current catch and maximum catch. Such approaches in other guises are similarly flawed^{12–15}.

Methods

Sustainable yield from fisheries as a function of population size can be represented by a Pella–Tomlinson curve¹⁶. From this curve, food provision is the ratio of sustainable yield to MSY, and can be written as a function of current biomass (B) relative to biomass producing MSY (B_{MSY}) plus a parameter (z) that governs the shape of the curve:

food provision =

$$100 \times \max \left\{ 0, \frac{z^{z/(z-1)}}{z-1} \left(\left(\frac{1}{z} \right)^{1/(z-1)} \frac{B}{B_{\text{MSY}}} - \left[\left(\frac{1}{z} \right)^{1/(z-1)} \frac{B}{B_{\text{MSY}}} \right]^z \right) \right\}$$

Setting $z = 1.188$ ensures that B_{MSY} occurs at 40% of unexploited biomass, which is the mean from 147 fisheries¹⁶. This food provision index is 0% when biomass is zero or greater than $2.5 \times B_{\text{MSY}}$, and 100% when biomass equals B_{MSY} .

This project was partly funded by Cooperative Agreement NA09OAR4320075 between the Joint Institute for Marine and Atmospheric Research (JIMAR) and the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not necessarily reflect the views of the NOAA or any of its subdivisions.

Trevor A. Branch¹, Daniel J. Hively¹ & Ray Hilborn¹

¹School of Aquatic & Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195, USA.
email: tbranch@uw.edu

Received 31 October 2012; accepted 28 January 2013.

1. Halpern, B. S. *et al.* An index to assess the health and benefits of the global ocean. *Nature* **488**, 615–620 (2012).
2. Pauly, D. *Focusing One's Microscope: Daniel Pauly on the State of Global Fisheries* <http://blog.nature.org/2011/01/daniel-pauly-fish-stock-global-world-fisheries/> (2011).
3. Worm, B. *et al.* Rebuilding global fisheries. *Science* **325**, 578–585 (2009).
4. Froese, R. & Proelß, A. Rebuilding fish stocks no later than 2015: will Europe meet the deadline? *Fish Fish.* **11**, 194–202 (2010).
5. Sumaila, U. R. *et al.* Benefits of rebuilding global marine fisheries outweigh costs. *PLoS ONE* **7**, e40542 (2012).
6. Arnason, R., Kelleher, K. & Willmann, R. *The Sunken Billions. The Economic Justification for Fisheries Reform.* (World Bank and FAO, 2008).
7. Costello, C. *et al.* Status and solutions for the world's unassessed fisheries. *Science* **338**, 517–520 (2012).
8. Ryther, J. Photosynthesis and fish production in the sea. *Science* **166**, 72–76 (1969).
9. Pauly, D. One hundred million tonnes of fish, and fisheries research. *Fish. Res.* **25**, 25–38 (1996).
10. FAO. *The State of World Fisheries and Aquaculture 2012* (Food and Agriculture Organization of the United Nations, 2012).
11. Ricard, D., Minto, C., Baum, J. K. & Jensen, O. P. Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish Fish.* **13**, 380–398 (2012).
12. Branch, T. A., Jensen, O. P., Ricard, D., Ye, Y. & Hilborn, R. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. *Conserv. Biol.* **25**, 777–786 (2011).
13. Daan, N., Gislason, H., Pope, J. G. & Rice, J. C. Apocalypse in world fisheries? The reports of their death are greatly exaggerated. *ICES J. Mar. Sci.* **68**, 1375–1378 (2011).
14. Carruthers, T. R., Walters, C. J. & McAllister, M. K. Evaluating methods that classify fisheries stock status using only fisheries catch data. *Fish. Res.* **119–120**, 66–79 (2012).
15. Hilborn, R. & Branch, T. A. Does catch reflect abundance? No, it is misleading. *Nature* **494**, 303–306 (2013).
16. Thorson, J. T., Cope, J. M., Branch, T. A. & Jensen, O. P. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. *Can. J. Fish. Aquat. Sci.* **69**, 1556–1568 (2012).

Author Contributions T.A.B. designed the study, T.A.B. and D.J.H. analysed the data, and T.A.B. and R.H. discussed the results and wrote the paper together.

Competing Financial Interests Declared none.

doi:10.1038/nature11974

Halpern *et al.* replyREPLYING TO T. A. Branch, D. J. Hively & R. Hilborn. *Nature* **495**, <http://dx.doi.org/10.1038/nature11974> (2013)

Branch *et al.*¹ suggest that the fisheries component of our ocean health and benefits index² is 'biased'. We contend that our approach is no more biased than their alternative, and that our method improves on theirs in three fundamental ways: it provides a score for every country, accounts for thousands of data-poor stocks, and incorporates sustainability more comprehensively.

Our global score is an area-weighted country average of a multi-dimensional assessment of sustainable food provision, fundamentally different from what Branch *et al.*¹ interpret it to be, which undermines comparisons to their global estimate. They incorrectly state that our index should be roughly equivalent to dividing global catch by global maximum sustainable yield (MSY), and incorrectly interpret a score of 100 in our index as $B = B_{\text{MSY}}$. They focus on single-stock biomass harvested relative to single-stock MSY, as in their Fig. 1, and assume that they know true MSY values. We used a multispecies reference point (mMSY) with a precautionary buffer that penalizes countries both above and below this value, and our focus was on fisheries within exclusive economic zone (EEZ) scales. Branch *et al.*¹ focused on a single global assessment (of current status only), leading to a relatively small number of the largest fish stocks worldwide driving their interpretation. We chose to use catch data, available for all countries and most stocks, acknowledging the associated limitations and uncertainty of those data. Although their approach may be useful for assessments of global biomass production, and we agree it is preferable to use abundance estimates and other indicators rather than catch data when possible, reliance on such data precludes country-level assessments at the global scale.

By focusing their analysis on a geographically restricted data set of exclusively large commercial species (<5% of all stocks in the world), they assume that small stocks, and the many interactions they have with large stocks, do not matter for understanding sustainability, and that the many data-poor stocks are irrelevant to understanding food provision. Data-poor stocks represent ~80% of all fisheries biomass and the vast majority of stocks in most countries. These stocks are in worse condition than assessed stocks, and getting worse³. Their selection of tuna and billfish as a counter example to our results focuses on an outlier; these large pelagic stocks are not representative of most fisheries within EEZs (that is, bottom-dwelling species or those taken in small-scale fisheries) or most unassessed stocks. Therefore, it is reasonable to assume that unassessed stocks would lower fisheries index scores, and for our comprehensive assessment to differ from the narrower one of Branch *et al.*¹

Because of these key differences, strong correlations between our methods should not necessarily be expected. In addition, assessments

made using best available data in individual countries should generate different outcomes compared with a global assessment where the focus is on using globally available and comparable data. We know that many country-specific scores will change as we improve the underlying data and our ability to predict sustainable yield of stocks around the world. Our index was explicitly designed as a framework that can ingest the best scientific information as it becomes available, whether as better data or improved approaches to measuring a goal, and in fact, we are doing just that as we apply the index at regional scales. Authors from both papers are currently collaborating with the Food and Agriculture Organization of the United Nations to examine how to improve methodologies for estimating fisheries exploitation status globally. We look forward to incorporating those results and improved data in future iterations of the index.

This Reply is written by the subset of authors that developed the fisheries component of the ocean health and benefits index² relevant to the discussion in Branch *et al.*¹.

Benjamin S. Halpern^{1,2}, Steven D. Gaines³, Kristin Kleisner⁴, Catherine Longo¹, Daniel Pauly⁴, Andrew A. Rosenberg⁵, Jameal F. Samhouri⁶ & Dirk Zeller⁴

¹National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA.

²Center for Marine Assessment and Planning, University of California, Santa Barbara, California 93106, USA.
email: halpern@nceas.ucsb.edu

³Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, USA.

⁴Sea Around Us Project, Fisheries Centre, 2202 Main Hall, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

⁵Union of Concerned Scientists, 2 Brattle Square, Cambridge, Massachusetts 02138, USA.

⁶Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, Washington 98112-2097, USA.

1. Branch, T. A., Hively, D. J. & Hilborn, R. Is the ocean food provision index biased? *Nature* **495**, <http://dx.doi.org/10.1038/nature11974> (2013).
2. Halpern, B. S. *et al.* An index to assess the health and benefits of the global ocean. *Nature* **488**, 615–620 (2012).
3. Costello, C. *et al.* Status and solutions for the world's unassessed fisheries. *Science* **338**, 517–520 (2012).

doi:10.1038/nature11975